

VANIER

C É G E P / C O L L E G E

ALIGNEMENT CURRICULAIRE EN FRANÇAIS LANGUE SECONDE AU CÉGEP VANIER COURS DE NIVEAU 1



**PHILIPPE GAGNÉ
ISABELLE DER APRAHAMIAN
LAURENCE DUMONT**

**LA PRÉSENTE RECHERCHE A ÉTÉ SUBVENTIONNÉE
PAR LE CÉGEP VANIER DANS LE CADRE DES *ACADEMIC SUCCESS PROJECTS***

**LE CONTENU DU PRÉSENT RAPPORT N'ENGAGE
QUE LA RESPONSABILITÉ DE L'ÉTABLISSEMENT ET DE L'ÉQUIPE DE RECHERCHE**

JUIN 2015

Conception graphique de la couverture : David Spadotto

© 2015, Vanier College Press
Tous droits réservés

Le présent rapport a été rédigé en suivant les principes généraux de la rédaction épiciène de l'Office québécois de la langue française afin d'accorder une visibilité équivalente aux genres féminin et masculin. Il est aussi conforme à la nouvelle orthographe.

La reproduction d'extraits de cet ouvrage est autorisée avec mention de la source.

Il est possible de télécharger ce rapport en format PDF à l'adresse suivante :
<http://www.cdc.qc.ca/pdf/788739-gagne-et-al-alignement-curriculaire-FLS-vanier-2015.pdf>

Pour tout renseignement sur ce rapport, s'adresser à :

Philippe Gagné
821, av. Ste-Croix, bur. K-327
Montréal (Québec) H4L 3X9
gagnep@vaniercollege.qc.ca

ISBN : 978-2-923990-10-1 (imprimé)
978-2-923990-11-8 (PDF)

Dépôt légal — Bibliothèque et Archives nationales du Québec, 2015

Descripteurs : Français langue seconde, cégep, enseignement supérieur, alignement curriculaire, évaluation, équité, validité, lisibilité

REMERCIEMENTS

Nous remercions nos collègues du Département de français langue seconde du cégep Vanier pour la confiance qu'ils nous ont manifestée en acceptant de participer à cette étude. Nous remercions aussi les membres du bureau de la Direction des études de Vanier pour leur soutien financier et moral.

Nous devons vivre, sans aucun doute, avec l'idée qu'il y a de l'arbitraire dans nos évaluations. Par contre, pour être à la hauteur de la confiance qu'on nous accorde comme professionnels, il me semble qu'on doit poursuivre inlassablement l'effort d'améliorer nos pratiques évaluatives. Dans la mesure où un enseignant est conscient des enjeux éthiques de l'évaluation et cherche avec ses collègues à trouver des solutions, il fait déjà preuve de professionnalisme.

S'il y a une part d'arbitraire, il y a aussi une part d'objectivité dans les évaluations. Il faut alors miser sur les stratégies qui pourront augmenter cette part d'objectivité.

Denis Jeffrey, Université Laval, (2009)

Table des matières

Liste des figures et des tableaux.....	7
RÉSUMÉ	8
1. INTRODUCTION	8
2. PROBLÉMATIQUE.....	9
2.1. L'application de la politique institutionnelle d'évaluation des apprentissages (PIEA)....	10
2.2. De l'approche commune à l'approche individuelle	13
2.3. Énoncé du problème.....	16
2.4. Résultats escomptés	17
2.5. Pertinence de la recherche	17
3. CADRE THÉORIQUE.....	18
3.1. Les habiletés de réception : la compréhension de l'écrit et de l'oral	22
3.2. Les habiletés de production : l'expression écrite et orale	28
3.3. Définition opérationnelle de l'alignement curriculaire.....	29
4. MÉTHODOLOGIE.....	29
5. RÉSULTATS.....	32
5.1. Données descriptives globales.....	32
5.2. Description des items	33
5.3. La compréhension écrite	39
5.4. Le niveau de difficulté des textes à lire.....	41
5.5. La compréhension orale	46
5.6. L'expression écrite	48
5.7. L'expression orale	49
5.8. Les types de connaissances.....	51
5.9. Synthèse et conclusion	52
RÉFÉRENCES.....	56

Liste des figures et des tableaux

Figure 1 Connaissance du lexique français selon la fréquence des mots dans la langue par tranche de 1000 mots au niveau 1	25
Figure 2 Schéma illustrant le processus d'une analyse d'alignement curriculaire au sein d'un département	31
Figure 3 Proportion d'items par objectif ministériel.....	34
Figure 4 Valeur attribuée par objectif ministériel.....	35
Figure 5 Valeur attribuée à chaque type de question.	38
Figure 6 Proportion d'items par catégorie RQR en compréhension de l'écrit	40
Figure 7 Valeur dans le cours de chaque catégorie RQR en compréhension de l'écrit	41
Figure 8 Proportion d'items par catégorie RQR en compréhension de l'oral	47
Figure 9 Valeur dans le cours de chaque catégorie RQR en compréhension de l'oral	48
Figure 10 Proportion d'évaluations par catégorie en expression écrite	49
Figure 11 Proportion d'évaluations par catégorie en expression orale	50
Figure 12 Proportion du nombre d'items selon le type de connaissances	51
Figure 13 Valeur attribuée par type de connaissances.....	52

Tableau 1 Comparaison des recommandations de la CEEC (2011) avec les pratiques du département de FLS.....	12
Tableau 2 Barème de notation commun pour les cours de FLS au niveau 1	14
Tableau 3 Résultats à l'épreuve de vocabulaire, au niveau 1, dans l'étude de Michaud (2014) ..	25
Tableau 4 Taxonomie pour les évaluations de l'expression écrite	28
Tableau 5 Taxonomie pour les évaluations de l'expression orale	29
Tableau 6 Description du contenu des quatre cours étudiés	33
Tableau 7 Nombre d'items par objectif ministériel	34
Tableau 8 Valeur attribuée par objectif ministériel.....	35
Tableau 9 Nombre d'items par type de question pour chaque objectif ministériel	37
Tableau 10 Proportion d'évaluations selon le format	39
Tableau 11 Nombre d'items par catégorie RQR en compréhension de l'écrit.....	40
Tableau 12 Tableau synthèse des caractéristiques superficielles des textes à lire en fin de session au niveau 1.....	46
Tableau 13 Nombre d'items par catégorie RQR en compréhension de l'oral.....	47
Tableau 14 Nombre d'items par type de connaissance	51

RÉSUMÉ

Ce rapport présente une analyse de l'alignement curriculaire de quatre cours différents de niveau 1 (602-100-MQ) donnés par quatre membres du département de français langue seconde au cégep Vanier. Le but est, d'une part, de mesurer la congruence entre les évaluations et les objectifs ministériels correspondants et, d'autre part, de mesurer le degré d'équité entre les cours. En ce sens, les items d'évaluation en expression écrite et orale de même qu'en compréhension écrite et orale ont été codés au moyen de taxonomies choisies afin de raffiner la connaissance de ce qui est évalué. Les résultats montrent que les cours évaluent tous les objectifs ministériels, mais dans des proportions différentes. Lorsque confrontés aux mesures taxonomiques, les cours présentent certaines disparités. Il en va de même pour les modalités d'évaluation en classe. De plus, il est démontré que le barème de notation commun échoue à garantir l'équivalence des évaluations entre les cours. Enfin, un élément original de cette étude consiste à proposer une démarche pour calibrer les textes au programme (lecture et écoute) de façon à favoriser l'équivalence sur le plan du niveau de difficulté.

This report presents a curriculum alignment analysis of four different level 1 courses (602-100-MQ) taught by four French as a second language teachers at Vanier College. The aim is, firstly, to measure the congruence between assessment objectives and corresponding ministerial standards and, secondly, to measure the degree of equity between sections of the same course. In this sense, assessment items in written and oral expression as well as in reading and listening comprehension were coded using selected taxonomies to refine the understanding of what is being evaluated. The results show that the courses assess all ministerial objectives but in different proportions. When analyzed from the perspective of taxonomic categories, the various sections of the same course reveal differences, which have been highlighted. The same goes for classroom assessment procedures. Moreover, it is shown that the common marking scheme fails to guarantee the equivalence of assessments between sections. Finally, an original feature of this study is to propose an approach to calibrate the texts (reading and listening) so as to promote equivalence in terms of level of difficulty.

1. INTRODUCTION

Cette étude fait suite à une autre étude similaire effectuée en 2012 auprès de collègues donnant des cours au niveau 3 en français langue seconde (FLS) au cégep Vanier (Gagné, Brunet, Boucher et Dumont, 2012). Elle vise à mesurer le degré de cohérence ou la comparabilité de l'évaluation dans quatre cours de niveau 1 donnés par plusieurs enseignantes et enseignants avec des contenus différents. Est-ce que les élèves ont des opportunités d'apprentissage comparables d'un cours à l'autre? Est-ce que les évaluations de ces cours sont équivalentes à partir de points de comparaison comme les types de connaissances invoquées

dans les tâches à réaliser? Est-ce que les questions de compréhension en lecture et en écoute sont comparables? Est-ce que le niveau de difficulté des textes à lire est comparable d'un cours à l'autre? Ces questions font partie des enjeux auxquels le corps enseignant fait face puisqu'il s'agit d'une responsabilité départementale inscrite dans la politique institutionnelle d'évaluation des apprentissages (PIEA) du Collège. L'analyse de l'alignement curriculaire, aussi appelée cohérence curriculaire, est un moyen pour vérifier cette comparabilité entre les cours. Ce type d'analyse est rare dans le réseau collégial, mais le Collège régional Champlain Saint-Lambert l'a intégrée à ses pratiques d'évaluation institutionnelle dans la majorité de ses programmes et départements (Bateman, Taylor, Janik et Logan, 2009). Le présent rapport s'appuie sur cette expertise et l'adapte à ses besoins disciplinaires particuliers.

Les deux premiers coauteurs de ce rapport sont des membres du département de FLS. Ils sont les experts de contenu qui ont effectué le travail de collecte des données, d'analyse et de rédaction du rapport. Une ressource externe, la dernière coauteure, est une doctorante en psychologie à l'Université de Montréal. Elle a effectué le traitement informatique et statistique des données en plus de collaborer à leur analyse et à la rédaction du rapport.

2. PROBLÉMATIQUE

Depuis la réforme Robillard de 1993, la discipline de FLS fait partie de la formation générale obligatoire du secteur anglophone des cégeps québécois – avec la littérature anglaise, les humanités, et l'éducation physique – et appuie la formation spécifique pour favoriser l'obtention d'un Diplôme d'études collégiales (DEC) dans un programme préuniversitaire ou technique. Tous les candidats au DEC doivent suivre deux cours obligatoires de 45 heures en FLS, le premier dit de la « formation commune » et le second de la formation « propre au programme », un cours de français sur objectifs spécifiques (MELS, 2011). Les quatre habiletés langagières y sont enseignées : lire, écrire, parler et écouter, sauf dans les niveaux intermédiaire et avancé où l'accent est mis sur la lecture et l'écriture. Le département de FLS du cégep Vanier, un collège public de 6500 élèves, compte bon an mal an entre 20 et 25 membres. Il donne trois cours de la formation spécifique et quinze cours à la formation ordinaire répartis sur quatre

niveaux de compétence (faible à avancé, les cours 602-100-MQ à 602-103-MQ). Les niveaux débutants sont dispensés à l'Éducation permanente. Un cours de Mise à niveau (602-009-50) pour les élèves ayant réussi la 5^e secondaire avec 65 % ou moins dans le programme de base de FLS et pour les élèves arrivés récemment de l'extérieur du Québec est aussi offert. Chaque session, près de 70 groupes de 35 élèves sont constitués pour suivre l'un ou l'autre de ces cours. Seuls trois cours de la formation spécifique sont donnés par un seul membre du département alors que les autres sont donnés simultanément par cinq à huit membres en général. C'est-à-dire qu'un même cours sera donné en cinq à huit versions différentes avec des contenus, des thèmes, des évaluations et des critères d'évaluation différents. Par contre, ces cours ont en commun un barème de notation (ex. : 15 % à la composition finale, etc.), des dates d'évaluation similaires, des balises à l'égard du nombre de mots à lire, de la durée des enregistrements à écouter, de la longueur des textes à écrire, de même que de leur genre.

2.1. L'application de la politique institutionnelle d'évaluation des apprentissages (PIEA)

En 2011, la Commission d'évaluation de l'enseignement collégial (CEEC) a remis un rapport d'*Évaluation de l'application de la politique institutionnelle d'évaluation des apprentissages du Collège Vanier* (CEEC, 2011 : 6-15) qui observait « que des améliorations sont à apporter, notamment par rapport à (...) l'équivalence des évaluations ». Dans cette portion de l'analyse, la Commission indique que des membres du corps enseignant « ont témoigné se concerter pour maintenir une similarité dans la charge de travail, le type de travaux demandés aux étudiants, les critères de correction ou le niveau de difficulté des évaluations sommatives. Cette collaboration se concrétise, dans certains départements, par des plans de cours ou des examens communs (...) » (p. 8). De plus, « pour vérifier que les modes et les instruments d'évaluation sont justes, équitables et équivalents, dans le cas d'un cours dispensé par plus d'un professeur, [le Collège] doit notamment mettre en place un comité (*curriculum committee*) chargé d'établir les modalités permettant d'assurer que les cours et les activités d'évaluation sont examinés régulièrement. À la lumière des documents qu'elle a consultés et des propos des divers intervenants rencontrés en visite, la Commission constate que cette vérification n'est pas effectuée » (p. 9). La Commission a donc « suggéré » — l'équivalent d'une recommandation formelle — au Collège de « s'assurer que les départements (...) exercent leurs responsabilités

relatives aux modes et aux instruments d'évaluation comme stipulé dans la PIEA » (p. 9). Enfin, lorsqu'un cours est donné par plus d'un membre du corps enseignant, les observations de la Commission l'ont amenée à suggérer « au Collège de s'assurer de l'équivalence de l'évaluation » (p. 15) qui « repose (...) sur la nécessaire concertation entre les professeurs » (Beumier, 2012 : 17). La Commission conclut que « l'application que le Collège Vanier a faite de sa PIEA n'assure pas la qualité de l'évaluation des apprentissages » (CEEC, 2011 : 18).

Étant donné que ce rapport est global pour l'ensemble du Collège, il est difficile de savoir quelles observations ou suggestions s'appliquent au département de FLS. Par contre, le tableau 1 souligne le fait que les membres du département 1) ne se concertent pas pour maintenir un niveau de difficulté similaire dans les évaluations sommatives ni une similarité dans la charge de travail — le nombre de tâches d'évaluation peut aller du simple à près du double dans les cours du niveau 3 (Gagné, Brunet, Boucher et Dumont, 2012 : 11) —, 2) qu'ils le font pour le type de travaux demandés aux élèves et 3) qu'ils n'ont pas de critères de correction équivalents pour l'évaluation de l'écriture et de l'expression orale. De plus, cette faible concertation ne débouche pas, comme le décrit la CEEC à propos des enseignantes et des enseignants qui ont témoigné se concerter, sur l'existence d'examens communs dans les cours donnés par plus d'un collègue. Il est aussi possible d'affirmer que le comité de programme (*curriculum committee*) du département n'examine pas régulièrement les cours et les activités d'évaluation comme stipulé dans la PIEA. Enfin, une expérience non scientifique menée en 2013 au sein du département a montré qu'il y avait un problème de fidélité des jugements de l'expression écrite. En effet, une dizaine de collègues se sont prêtés à un exercice de correction d'une même rédaction avec leur propre grille d'évaluation. Le résultat a montré une variation dans les notes attribuées allant de 57 % à 86 % pour cette rédaction. Une hypothèse pour expliquer ce résultat s'appuie sur le fait que les critères d'évaluation et le mode de distribution des notes étant différents pour tous les collègues, ces derniers n'évaluaient pas la même chose ni ne distribuaient les notes de la même

La Commission conclut que « l'application que le Collège Vanier a faite de sa PIEA n'assure pas la qualité de l'évaluation des apprentissages ».

manière. Cette expérience ne faisait que reproduire ce qui avait déjà été démontré à maintes reprises en éducation (Jeffrey, 2009; Roegiers, 2010).

Tableau 1 Comparaison des recommandations de la CEEC (2011) avec les pratiques du département de FLS

Recommandations de la CEEC adaptées à la discipline	Pratiques départementales
Concertation pour maintenir un niveau de difficulté similaire dans les évaluations sommatives	Non.
Concertation pour assurer une charge de travail similaire	Non.
Concertation sur le type de travaux demandés	Partiellement. La catégorie « contrôles et travaux pratiques » (30 %) est laissée à la discrétion des membres (voir tableau 2).
Critères de correction équivalents pour l'évaluation de l'expression écrite et de l'expression orale	Non.
Examens communs dans les cours donnés par plus d'un professeur	Non.
Examen des cours et des évaluations par le comité de programme (<i>curriculum committee</i>)	Partiellement. Les descriptifs de cours d'environ 100 mots et une médiagraphie sont approuvés par le comité de programme.

Si l'application de la PIEA pose problème à Vanier, la politique elle-même est adéquate. En 2002, la CEEC l'a évaluée et a souligné « la pertinence [de son] contenu (...) et elle [a] constat[é] une nette amélioration dans la présentation de ses composantes » (CEEC, 2002 : 4). La PIEA énonce les règles suivantes :

Le Collège confie aux départements et programmes la supervision des modes d'évaluation dans les cours. Ces mesures peuvent comprendre des textes communs, des critères de notation communs, des devoirs et des examens communs pour le même cours ou des cours similaires enseignés par des membres du département différents. Lorsque des mesures communes ne sont pas en place, le département, le programme ou le comité du cours détermine les principes relatifs à la conception, à la pondération et à l'évaluation des instruments de mesure utilisés pour évaluer l'apprentissage des élèves.

Chaque département doit vérifier que toutes les tâches d'évaluation et tous les modes d'évaluation décrits par l'enseignante ou l'enseignant sont justes, respectent les standards partagés et sont équitables pour les élèves de tous les groupes.

C'est la responsabilité des départements, de la faculté et des programmes d'implanter les modes d'évaluation des apprentissages des élèves.

Tout le corps enseignant a les responsabilités suivantes : (...) collaborer avec les collègues quand ils enseignent le même cours à des groupes différents afin de maintenir la cohérence et la fidélité (Traduction libre, désormais TL, de Vanier College, 2011).

2.2. De l'approche commune à l'approche individuelle

À la création des cours obligatoires, en 1994, le département de FLS a choisi l'approche des « mesures communes » pour respecter ces directives, sauf dans le niveau 4 (602-103-MQ), le niveau avancé, où le choix des œuvres et des pratiques pédagogiques et d'évaluation était laissé à la discrétion des membres. En effet, dans tous les autres niveaux, tous les membres donnaient à peu de choses près le même cours avec des évaluations communes. À l'arrivée d'une nouvelle vague de collègues, vers le milieu des années 2000, ces derniers ont exprimé la volonté de créer de nouveaux cours personnalisés selon les intérêts de chacun au nom d'une certaine liberté d'action. L'assemblée départementale a accepté ce changement de cap en 2009 et un élan de créativité a permis la conception de dizaines de cours à tous les niveaux de compétence. Toutefois, la responsabilité que le Collège confie au corps enseignant d'assurer l'équivalence des évaluations dans le cas des cours donnés par plus d'un membre du département demeurait, mais seules quelques mesures déjà présentes au département sont restées en place : un barème de notation commun (par exemple, la valeur des évaluations de lecture dans le cours, etc., voir tableau 2) et les prescriptions présentes dans le devis ministériel à propos de la longueur des textes à écrire et à lire, la durée des exposés oraux, etc. (MELS, 2011). Comme indiqué ci-dessus, le choix des évaluations de la catégorie « contrôles et travaux pratiques » est laissé à la discrétion des membres du département. Aucune mesure de l'adéquation entre les évaluations et les objectifs ministériels ou pour déterminer la comparabilité des évaluations des différents cours n'a été instaurée pour « maintenir la cohérence et la fidélité ».

Tableau 2 Barème de notation commun pour les cours de FLS au niveau 1

Évaluations	Valeur dans le cours
Contrôles et travaux pratiques	30 %
Examen de compréhension de l'écrit 1	10 %
Examen de compréhension de l'écrit 2	15 %
Examen de compréhension de l'oral 1	5 %
Examen de compréhension de l'oral 2	5 %
Expression écrite	25 %
Présentation orale	10 %

La réalité d'un département de cégep fait en sorte que de telles mesures ne sont pas faciles à implanter. Chaque membre jouit d'une grande autonomie et n'a pas de supervision dans son travail de la part d'un conseiller pédagogique ou d'un directeur comme c'est le cas au primaire et au secondaire. En effet, la Commission scolaire de Montréal, par exemple, impose des situations d'évaluation validées (épreuves développées à l'interne ou avec des gens d'autres commissions scolaires). Dans un second temps, elle procède à une collecte d'échantillonnage en demandant aux membres du corps enseignant de lui envoyer des productions d'élèves (6^e année et 2^e secondaire) sélectionnées au hasard pour « [s'assurer] que les élèves apprennent bien ce qu'ils doivent apprendre et pour identifier les forces et les défis des élèves afin de bien planifier les formations et accompagnements pour l'année suivante » (Germain, 2015). Au collégial, le département est le centre décisionnel à l'égard des questions d'ordre pédagogique et didactique, mais le degré de collaboration entre ses membres sur les éléments étudiés par la CEEC est tenu en général. Lorsque les membres sont invités à mettre en place des pratiques visant à assurer la cohérence curriculaire¹ des cours, une certaine résistance s'affirme au nom de l'autonomie professionnelle ou de la « liberté académique ». À ce sujet, il est intéressant de lire la position des principaux syndicats du corps enseignant à l'égard de la définition du concept d'autonomie individuelle, des fonctions de l'assemblée départementale de même que de la responsabilité de la coordination départementale envers le Collège :

L'autonomie individuelle (...) est imbriquée dans l'exercice d'une autonomie collective qui en détermine le contour et [est] balisée par celui-ci. (p. 1.2.4)

¹ « L'alignement curriculaire vise à assurer une correspondance élevée entre le programme prescrit, l'enseignement réalisé et l'évaluation effectuée (Cohen, 1987 et 1995; Guskey, 2003) » (Gauthier et collab., 2013 : 28). Ce concept sera approfondi dans le cadre théorique.

L'autonomie n'est toutefois pas l'indépendance et les choix ne peuvent être uniquement individuels. Ils doivent se faire dans un cadre partagé et assumé par un ou par des collectifs professionnels. Le lieu premier de cette régulation, c'est le département. (p. 1.5)

Fonctions de l'assemblée départementale (article 4-1.05) : (...) définir les objectifs, appliquer les méthodes pédagogiques et établir les modes d'évaluation spécifiques à chacun des cours dont elle est responsable en tenant compte de la Politique institutionnelle d'évaluation des apprentissages; (...)

Activités dont le coordonnateur rend compte au collège : (...) s'assurer que soient définis les objectifs, appliquées les méthodes pédagogiques et établis les modes d'évaluation propres à chacun des cours dont le département est responsable en tenant compte de la Politique institutionnelle d'évaluation des apprentissages (p. 1.3.4 et 1.3.5) (Sabourin, 2009).

En somme, il revient au département d'assumer les responsabilités qui lui incombent de par la PIEA et au coordonnateur de se porter garant de la réalisation des tâches liées à ces responsabilités : évaluer les instruments de mesure, s'assurer qu'ils sont équitables pour les élèves de tous les groupes et veiller à ce que les membres du département collaborent lorsqu'ils enseignent le même cours à des groupes différents afin de maintenir la cohérence et la fidélité. L'approche commune favorisait le respect de ces responsabilités, mais l'approche individuelle ne permet pas de le favoriser puisqu'aucun mécanisme d'évaluation n'a été mis en place pour vérifier si toutes les versions d'un même cours étaient cohérentes les unes avec les autres (cohérence horizontale) et avec les devis ministériels (cohérence verticale).

L'approche commune favorisait le respect de ces responsabilités, mais l'approche individuelle ne permet pas de le favoriser puisqu'aucun mécanisme d'évaluation n'a été mis en place pour vérifier si toutes les versions d'un même cours étaient cohérentes les unes avec les autres et avec les devis ministériels.

La présente étude de cohérence curriculaire constitue un effort pour pallier ce manque tout comme celle réalisée pour les cours du niveau 3 citée précédemment. Elle s'inspire des travaux effectués par Bateman et son équipe (2009) au Collège Champlain Saint-Lambert. Ces travaux servent de modèle puisque cet établissement a fait de l'analyse de la cohérence des programmes et des départements un processus continu depuis près de 15 ans au moyen de l'alignement curriculaire. À ce propos, lorsque la CEEC a évalué son programme de Science, en 2006, elle a souligné le rôle déterminant joué par le groupe de Bateman dans la révision du programme et lui a attribué les « nettes améliorations » à l'égard de la cohérence des plans de

cours et du respect de la PIEA (CEEC, 2006 : 6). Au Québec, hormis ces travaux, les écrits en éducation font apparaître le concept d'alignement curriculaire en 2005 dans un rapport de la Chaire de recherche du Canada en formation à l'enseignement de l'Université Laval dont le contenu à ce sujet a été repris dans une monographie en 2013 (Gauthier et collab., 2013, 2005). C'est donc dire que cette pratique d'autoévaluation du programme prescrit et évalué est embryonnaire chez nous et qu'il n'est pas surprenant que sa mise en œuvre soit difficile au collégial. Toutefois, il serait souhaitable de l'instaurer avec la collaboration des membres afin que l'assemblée départementale de FLS puisse s'acquitter des responsabilités que lui a confiées le Collège dans sa PIEA. Loin d'être une menace à l'autonomie individuelle ou professionnelle,

Loin d'être une menace à l'autonomie individuelle ou professionnelle, l'alignement curriculaire est un moyen d'en donner davantage en évitant la contrainte des « mesures communes » tout en permettant d'assumer ses responsabilités face au Collège et aux élèves.

l'alignement curriculaire est un moyen d'en donner davantage en évitant la contrainte des « mesures communes » tout en permettant d'assumer ses responsabilités face au Collège et aux élèves.

2.3. Énoncé du problème

Le département de FLS du cégep Vanier a opté pour une approche individuelle dans la conception des cours donnés par plus d'un membre sans accompagner ce changement de mesures pour évaluer la cohérence entre les diverses versions d'un même cours. L'analyse d'alignement curriculaire est une mesure exemplaire pour pallier ce manque et permet de favoriser le respect des règles incluses dans la PIEA à cet égard. C'est pourquoi cette étude s'appuie sur ce mode d'analyse pour répondre à la question de recherche suivante :

À partir d'une analyse d'alignement curriculaire, quel est le degré de cohérence entre les évaluations des cours donnés par plusieurs membres du département au niveau 1 en FLS au cégep Vanier, soit deux de la formation commune et deux de la formation propre au programme? Cette recherche vise donc à établir le degré de cohérence entre les évaluations de quatre cours du niveau 1 en FLS au cégep Vanier au moyen d'une analyse d'alignement curriculaire.

2.4. Résultats escomptés

L'équipe de recherche effectue cette étude afin de rendre compte de ce degré de cohérence, d'une part, et, d'autre part, pour amener tous les collègues du département de FLS à participer à la réflexion que les résultats pourraient susciter et à appuyer les changements qui pourraient en découler par consensus. Pour l'instant, la première étude réalisée avec les cours du niveau 3 n'a pas mené à un appui de la démarche d'alignement curriculaire malgré les enjeux identifiés dans le rapport. Par exemple, un des principaux enjeux constitue la calibration des questions de compréhension de l'écrit en fonction de la relation entre la question posée et la réponse (RQR, (Giasson, 2011)). Les écarts entre les trois niveaux de la RQR (explicite et textuelle, implicite et textuelle et implicite et faisant appel aux connaissances antérieures) d'un cours à l'autre étaient notables et un standard a été proposé par l'équipe de recherche. Toutefois, l'application de ces pratiques pour répondre aux exigences de la PIEA a été rejetée par l'assemblée départementale au moment d'écrire ce rapport sans qu'une solution alternative soit proposée.

2.5. Pertinence de la recherche

Les modes d'application de l'alignement curriculaire sont bien connus aux États-Unis depuis une vingtaine d'années où le « *U.S. Department of Education's Joint Dissemination Review Panel* (Block et al., 1995) a identifié le principe de l'alignement curriculaire comme étant une stratégie d'innovation exemplaire, validée empiriquement et pouvant être recommandée pour diffusion au niveau national » (Gauthier et collab., 2005 : 33). Ils le sont beaucoup moins au Québec, mais Gauthier et ses collègues contribuent à les faire connaître. Leur traduction du concept de *curriculum alignment* est reprise ici, mais il apparaît que les connotations liées à l'« alignement » nuisent à la réception de sa valeur. Il serait sans doute plus juste de parler de cohérence curriculaire comme le fait Bateman, car c'est bien de ça qu'il est question et non d'un alignement de type militaire où chacun reproduit les mêmes gestes sous les ordres d'une hiérarchie dominante. Au contraire, l'esprit d'une telle approche vise à donner au corps enseignant la liberté de choisir les thèmes et les lectures abordés dans leur cours, par exemple, la liberté de concevoir leurs propres

Il serait sans doute plus juste de parler de cohérence curriculaire, car c'est bien de ça qu'il est question et non d'un alignement de type militaire où chacun reproduit les mêmes gestes sous les ordres d'une hiérarchie dominante.

évaluations en lien avec des standards établis en collégialité et le sentiment d'assumer les responsabilités professionnelles confiées par le Collège afin de rendre l'évaluation équitable pour tous les élèves, entre autres. De plus, une telle approche augmente la comparabilité des opportunités d'apprentissage dans chaque cours en veillant à ce que les évaluations témoignent de l'atteinte des objectifs ministériels du cours et offrent un niveau de difficulté similaire.

3. CADRE THÉORIQUE

Le *curriculum alignment* est appliqué et étudié aux États-Unis depuis une vingtaine d'années et joue un rôle plus important encore depuis la promulgation, en 2001, du *No Child Left Behind Act*, la loi qui ne laisse aucun enfant derrière (TL). L'objectif est d'arrimer les standards gouvernementaux avec les activités d'apprentissage et les modes d'évaluation pour « s'assurer que les élèves aient l'opportunité d'apprendre le contenu [prescrit] par les standards » (Squires, 2009 : 5). Cette pratique a pris une ampleur considérable aux États-Unis parce que le concept d'imputabilité, du corps enseignant, des écoles et des *districts*, a été exacerbé au point où le salaire des premiers et l'existence même des secondes dépendent des résultats obtenus par les élèves aux tests standardisés de l'État. Un si haut degré d'imputabilité a rendu crucial que les contenus enseignés et les évaluations soient en phase avec les standards et objectifs nationaux. Cependant, il a résulté des exigences de cette loi une situation préoccupante puisque l'objectif n'a progressivement plus été de favoriser l'atteinte des standards, mais la réussite aux tests standardisés, créant de nombreuses distorsions dans le système scolaire américain (Ravitch, 2010). Il n'en demeure pas moins que la recherche a montré la pertinence de l'approche et ses effets sur la réussite de même que sur le développement professionnel du corps enseignant qui approfondit son appropriation des objectifs et ses connaissances en mesure et évaluation (Martone et Sireci, 2009 : 1334). Ces deux auteurs rappellent les limites et les possibilités qu'offre cette démarche : « Même si nous ne pouvons ou ne devons pas évaluer tout ce qui est enseigné aux élèves, la recherche en alignement curriculaire peut montrer clairement jusqu'à quel point les standards ou les contenus pédagogiques ont été évalués ». Il s'agit de l'alignement vertical, celui entre les évaluations et les objectifs et standards

ministériels. L'alignement horizontal consiste à comparer les différentes versions d'un même cours enseigné par plus d'un membre du corps enseignant. Pour Webb (1997 : 4), l'alignement curriculaire est une méthode permettant de mesurer « le degré de concordance entre les attentes [les objectifs et standards] et les évaluations et jusqu'à quel point elles se complètent pour faire en sorte que les élèves apprennent ce que l'on s'attend qu'ils sachent et fassent » (TL).

Les deux principaux chefs de file en matière d'alignement curriculaire sont Marzano et Webb. Le premier s'est d'abord intéressé à la compilation des standards gouvernementaux américains des disciplines avec ceux des associations professionnelles nationales afin de les synthétiser en un seul document. Il a ensuite créé un site Web (www.achieve.org) contenant tous ces standards afin que les développeurs de tests ou de méthodes puissent y « aligner » leur matériel pédagogique. Pour ce faire, quatre critères servent de points de repère : le contenu, la performance (le test demande-t-il aux élèves de démontrer la maîtrise des habiletés prescrites?), le niveau de difficulté de même que l'équilibre et l'étendue (*balance and range*; l'ensemble des standards et objectifs est-il couvert? L'importance relative des objectifs est-elle bien calibrée dans les évaluations?) (Squires, 2009 : 50-51). Marzano demande à des experts de contenu de réviser les tâches de lecture pour s'assurer qu'elles mesurent bien la compréhension et non les connaissances antérieures (qu'il faut avoir lu le texte pour pouvoir répondre aux questions; les connaissances antérieures n'étant pas complètement exclues de ce qui permet aux élèves de répondre à des questions de compréhension, pour faire des inférences par exemple). Les experts révisent aussi les textes à lire et déterminent par consensus que le vocabulaire, la syntaxe, les aspects littéraires, l'intrigue, la structure sont adéquats pour le niveau. Enfin, ils évaluent les tâches d'écriture en fonction du vocabulaire, de la clarté de l'intention (*purpose*) et du destinataire. Les experts codent aussi le niveau d'exigence cognitive pour répondre à une tâche d'évaluation sur une échelle à quatre niveaux : 1) le rappel d'information ou une compréhension de base; 2) l'application d'une habileté ou d'un concept; 3) la pensée stratégique; 4) la réflexion profonde (*extended analysis*; généralement pendant une longue période de temps) (Martone et Sireci, 2009 : 1344-1345).

De son côté, Webb s'est intéressé au degré d'alignement entre les standards gouvernementaux américains des disciplines et les tests standardisés conçus pour en certifier l'atteinte. Pour y arriver, deux des approches qu'il a utilisées sont pertinentes pour le contexte collégial québécois. La première consiste en un développement séquentiel, c'est-à-dire que les standards et objectifs sont d'abord conçus et alignés entre eux (progression des compétences, etc.) et sont ensuite utilisés comme point de référence pour développer les seconds éléments, les programmes (*curriculum frameworks*) ou les évaluations. La seconde approche, la révision d'experts, fait appel à des expertes et à des experts de contenu, à l'instar de Marzano, pour réviser l'alignement entre les standards gouvernementaux et les évaluations comme le veut la tradition dans ce type d'étude (Martone et Sireci, 2009 : 1336; Squires, 2009). La première est à la base de l'élaboration de l'approche par compétences au collégial en 1993 et la seconde est celle que Bateman et ses collègues ont utilisée au Collège Champlain Saint-Lambert.

Webb applique quatre dimensions dans ses analyses d'alignement (Martone et Sireci, 2009 : 1338-1340) :

1. la concomitance des catégories (*Categorical Concurrence*), c'est-à-dire qu'il s'agit d'évaluer la présence d'éléments de contenu au sens large du terme (géométrie, le sens du nombre) et Webb suggère qu'un minimum de six items devraient mesurer un objectif ministériel pour certifier l'alignement;
2. les processus cognitifs (*Depth of Knowledge*), comparent le degré de complexité des processus cognitifs qu'exigent les objectifs ministériels avec le degré de complexité des mêmes processus qu'exigent les tâches d'évaluations (les mêmes processus cognitifs que Marzano). Le principal critère ici est qu'au moins 50 % des items d'évaluation devraient être au même niveau ou au-dessus du niveau de l'objectif auquel il correspond. Il s'agit ici d'une variable importante qui détermine si l'alignement vertical est réalisé.
3. l'envergure de la connaissance (*Range of Knowledge*) analyse le nombre d'objectifs d'un standard qui sont mesurés par au moins un item d'évaluation. Pour attester de l'alignement dans cette dimension, au moins 50 % des objectifs d'un standard doivent

être mesurés par au moins un item d'évaluation. Vu le petit nombre d'objectifs dans les devis ministériels au collégial, cette dimension n'est pas pertinente pour cette étude.

4. la représentation équilibrée (*Balance of Representation*) examine jusqu'à quel point les items sont distribués uniformément à travers les objectifs.

Une troisième voie, les *Surveys of Enacted Curriculum* (Porter et Smithson, 2001; Martone et Sireci, 2009 : 1347-1348), effectue le même type d'analyses que les deux premières, mais ajoutent la dimension de l'enseignement dans l'équation. C'est-à-dire que les membres de l'équipe de recherche sont appelés à coder le contenu pédagogique en se référant à une classe type, puis estiment l'importance donnée à chaque élément afin de déterminer leurs proportions sur le temps total d'enseignement. Bien que pertinente, cette méthode n'a pas été retenue dans cette étude pour des raisons de faisabilité.

Si le contexte québécois est différent en matière d'imputabilité, la méthode élaborée aux États-Unis peut servir à valider l'équivalence verticale et horizontale comme l'ont démontré Bateman et ses collègues (2009) dans leurs efforts pour augmenter la cohérence curriculaire. Cette équivalence s'appuie sur la cohérence interne entre les évaluations de cours différents ayant des visées d'évaluations semblables (par exemple, les évaluations sommatives de la compréhension de l'écrit ou de l'expression écrite des différentes versions d'un même cours), c'est-à-dire le « degré avec lequel les items ou les parties d'un instrument mesurent effectivement la même caractéristique ou le même objectif » (Legendre, 2005 : 229). Dans cette étude, la correspondance entre les évaluations et les objectifs ministériels est mesurée et les caractéristiques examinées sont : 1) la relation entre la question et la réponse (RQR) (Giasson, 2011; Irwin, 1986) pour la compréhension de l'écrit et de l'oral, 2) le niveau de difficulté des textes à lire lors de l'évaluation finale, 3) les éléments devant composer une tâche d'écriture et une tâche d'expression orale, 4) les différents types de tâches (choix multiples, phrases lacunaires, développement, etc.) et 5) les types de connaissance : déclaratives, procédurales et conditionnelles (Tardif, 1997).

Les processus cognitifs tels que définis par Marzano et Webb ou encore ceux utilisés par Bateman — la taxonomie de Bloom révisée par Anderson et Krathwohl (2001) — n'ont pas été

retenus dans cette étude. En effet, une importante limite propre à toutes les recherches qui examinent le degré de difficulté des évaluations ou des objectifs à partir de processus cognitifs a été établie dans la dernière version de *Educational Measurement* (Brennan, 2006), la référence dans ce domaine. « Des études ont montré que la dimension de la complexité cognitive est particulièrement difficile à évaluer (Le, 2002; Sanford et Fabrizio, 1999). (...) Même l'expert (*rater*) le mieux formé sera généralement incapable de déterminer sans ambiguïté quels processus sont requis par la plupart des items » (Koretz et Hamilton, 2006 : 556-557).

Enfin, les évaluations des cours ont été codées en fonction des types de connaissances sur lesquelles elles s'appuient.

Les **connaissances déclaratives** correspondent essentiellement à des connaissances théoriques, aux connaissances qui, à une certaine période, furent reconnues comme des savoirs. Il s'agit, selon Gagné (1985), de la connaissance de faits, de règles, de lois, de principes. (...) Exemples : la connaissance de la nécessité d'anticiper en lecture; l'énoncé des règles de l'accord des participes passés. (...) Les **connaissances procédurales** correspondent au comment de l'action, aux étapes pour réaliser une action, à la procédure permettant la réalisation de l'action. Fréquemment, dans les écrits pédagogiques, ces connaissances sont décrites comme des savoir-faire. (...) Exemples : lire un texte dont la structure est narrative; écrire un texte avec l'intention de persuader; accorder les participes passés dans un court paragraphe; anticiper la suite d'un texte à partir du premier paragraphe. (...) Les **connaissances conditionnelles** concernent le quand et le pourquoi. À quel moment et dans quel contexte est-il approprié d'utiliser telle ou telle stratégie, telle ou telle démarche, d'engager telle ou telle action? (...) Exemples : reconnaître un participe passé dans une phrase; reconnaître, entre deux textes, celui dont la structure est informative; distinguer, dans un texte, ce qui est essentiel et ce qui est secondaire. (Tardif, 1997 : 48-53)

3.1. Les habiletés de réception : la compréhension de l'écrit et de l'oral

En ce qui concerne les évaluations de la compréhension de l'écrit, la RQR (Pearson et Johnson, 1978) est retenue comme elle l'a déjà été dans l'étude portant sur les cours du niveau 3 (Gagné et collab., 2012). Cette taxonomie établit trois niveaux de relation :

1. **Explicite et textuel** : La relation est explicite et textuelle si la question et la réponse découlent toutes deux du texte et que la relation entre la question et la réponse est clairement indiquée par des indices dans le texte même.

Texte : Les flocons de neige qui tombent du ciel ressemblent à des étoiles blanches.
Question : À quoi ressemblent les flocons de neige qui tombent du ciel?
Réponse : Ils ressemblent à des étoiles blanches.

2. **Implicite et textuel** : La relation est implicite et textuelle si la question et la réponse découlent toutes deux du texte, mais qu'il n'y a pas dans le texte d'indice grammatical qui relie la question à la réponse. Cette catégorie demande l'utilisation d'au moins une inférence par le lecteur.

Texte : Dans la classe, nous avons deux poissons rouges, Pollux et Castor. Ce matin, il n'en reste plus qu'un. Pollux est mort.
Question : Castor est-il vivant ou mort?
Réponse : Castor est vivant.

3. **Relation implicite et faisant référence aux schémas du lecteur** : La relation est implicite et fondée sur les schémas du lecteur si la question seule découle du texte et que le lecteur utilise ses propres connaissances pour répondre à la question.

Texte : Ralph s'installa dans une vieille berceuse. Il se berça de plus en plus fort. Il se retrouva soudainement assis sur le plancher.
Question : Pourquoi Ralph se retrouva-t-il assis sur le plancher?
Réponse : Parce qu'une berceuse renverse quand on se balance trop fort. (Giasson, 2011).

Cette taxonomie a été choisie d'une part parce qu'elle fait partie d'un référentiel connu au Québec par l'entremise de l'ouvrage de Giasson et d'autre part parce qu'elle est similaire à d'autres taxonomies utilisées aux mêmes fins (Nation, 2009 : 34; Organisation de coopération et de développement économiques (OCDE), 2012 : 24).

Contrairement à Marzano, l'experte et l'expert de contenu de cette étude n'évaluent pas qualitativement le degré de lisibilité des textes au programme des cours analysés (v. infra). Ce type de travail correspond au 3^e type de connaissances clés reconnues pour établir le degré de difficulté d'un texte selon Murphy (2013) : 1) la connaissance des caractéristiques des lecteurs et de la tâche de lecture, 2) la connaissance des caractéristiques superficielles du texte, 3) la connaissance des caractéristiques profondes du texte et des modalités représentées dans le texte. Il s'agit d'une limite de cette étude puisqu'elle n'a pas les ressources pour couvrir les enjeux du 3^e élément qui exige une analyse de contenu au moyen d'un codage des textes par

catégories. Le 1^{er} élément sera abordé sous l'angle de la connaissance du vocabulaire français des élèves de ce niveau et d'une description de la tâche de lecture dans chaque cours étudié, puis le second type de connaissance sera traité quantitativement à l'aide de SATO-calibrage (Daoust, Laroche et Ouellet, 1996) et de VocabProfile (Selva, Jones, Goodfellow et Cobb, 2010), des instruments présents sur le Web.

Les élèves ne maîtrisent pas les 1000 mots les plus fréquents de la langue et cette « connaissance des caractéristiques des lecteurs » donne un éclairage nouveau sur la lecture à ce niveau.

La connaissance des caractéristiques des lecteurs et de la tâche de lecture est traitée en deux temps. D'une part, il est opportun de préciser que des élèves du niveau 1 à Vanier ont récemment fait l'objet d'une étude dans le cadre d'un mémoire de maîtrise et les résultats permettent de mieux cerner leurs capacités en lecture. D'autre part, la tâche de lecture de l'examen final de compréhension écrite est décrite

dans la section des résultats en fonction des modalités de passation : les élèves peuvent-ils lire les textes avant le jour de l'examen? ont-ils droit au dictionnaire? l'examen contient-il d'autres types de questions (grammaire, vocabulaire, etc.)? y a-t-il des questions sur les textes lus pendant la session?

L'étude mentionnée ci-dessus portant sur les caractéristiques des lecteurs visait à « étayer les connaissances que possèdent des apprenants de français langue seconde de niveaux intermédiaire [niveau 1] et intermédiaire avancé [niveau 2] en matière de traitement morphologique et de mettre en lien ces connaissances avec leurs connaissances en vocabulaire et leur compétence en compréhension écrite » (Michaud, 2014 : 79). Un test de connaissance du vocabulaire français en fonction de la fréquence du lexique dans la langue a été passé dans un groupe et les résultats sont présentés au tableau 3 et à la figure 1. Cet échantillon de 32 élèves a produit des résultats qui sont statistiquement valides. Avant de les observer, il est utile d'indiquer que, dans ses principes pour enseigner la lecture en langue seconde, Nation (2009 : 6, TL) indique que lorsque l'élève lit en focalisant sur le sens (*Meaning-focused Input*), les textes devraient être adaptés au niveau du groupe. En particulier, lorsque le contenu de la lecture vise à développer la maîtrise de la langue, les élèves devraient lire des textes dont

ils connaissent 98 % du vocabulaire « afin qu'ils puissent apprendre les 2 % restants en devinant à partir du contexte ». Les élèves qui ont participé à l'étude de Michaud ont montré une compréhension de 88,91 % des 1000 mots les plus fréquents et de 78,95 % de la tranche des 1001 à 2000 mots les plus fréquents de la langue française. Ce pourcentage continue de descendre au fur et à mesure que les mots sont plus rares. Ces résultats nous permettent de conclure que tous les élèves ne maîtrisent pas les 1000 mots les plus fréquents et cette « connaissance des caractéristiques des lecteurs » (Murphy, 2013) donne un éclairage nouveau sur l'enseignement et l'évaluation de la lecture à ce niveau.

Tableau 3 Résultats à l'épreuve de vocabulaire, au niveau 1, dans l'étude de Michaud (2014)

	N	Moyenne (%)	Écart type (%)	Hasard
Vocabulaire Total (/5000)	32	58,19	14,05	$p < 0,002$
1000 (/20)	32	88,91	12,75	-
2000 (/20)	32	78,95	15,65	-
3000 (/20)	32	69,20	18,62	-
4000 (/20)	32	62,34	21,33	-
5000 (/20)	32	45,47	14,05	-
Pseudomots	32	-	-	-

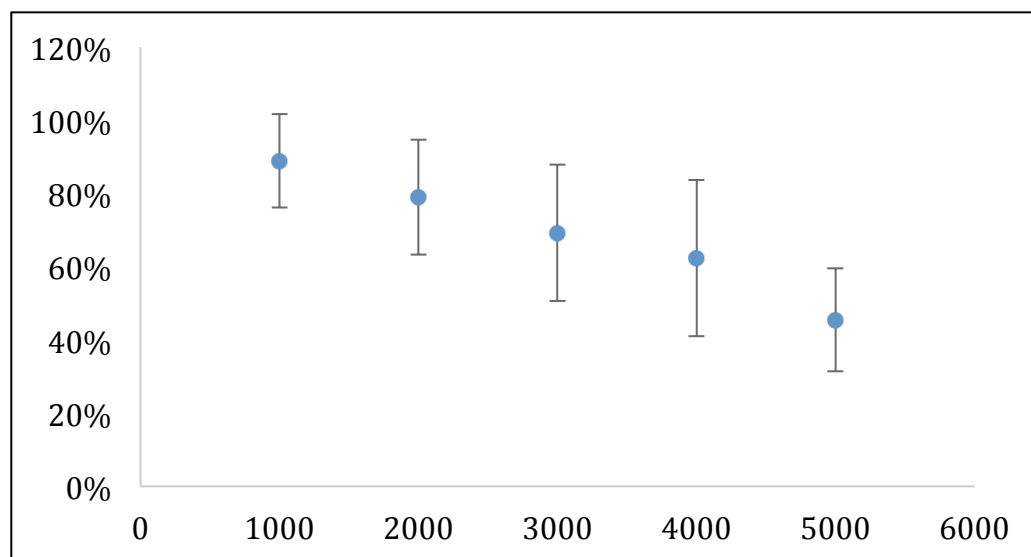


Figure 1 Connaissance du lexique français selon la fréquence des mots dans la langue par tranche de 1000 mots au niveau 1

La connaissance des caractéristiques superficielles du texte a été étudiée avec deux instruments informatiques. Le premier, SATO-calibrage, effectue une analyse fine du texte en examinant le nombre de syllabes par mots, de mots par phrase, de phrases par texte, le nombre de mots de neuf lettres ou plus, le nombre de relatives, etc. De plus, une base de données s'appuie sur un corpus de textes utilisés à tous les niveaux du primaire et du secondaire francophone québécois. Il est ainsi possible de calibrer un texte soumis à l'outil en le comparant à ce corpus qui agit comme un standard échelonnant les caractéristiques des textes par niveau scolaire de la 1^{re} à la 11^e année du cursus québécois. Par conséquent, une chercheuse ou un enseignant soumet un texte à l'outil en ligne qui indique à quel niveau scolaire du secteur francophone ce texte correspond. Le fait que cette donnée soit liée aux niveaux scolaires francophones constitue une limite à son utilisation. Toutefois, comme les cégépiens et les cégépiennes terminent leur enseignement obligatoire en FLS et qu'ils auront à travailler dans un milieu où le français devrait occuper une place importante à la fin de leurs études, il peut être intéressant de connaître cette correspondance avec l'enseignement de la langue maternelle après 11 ans de formation à Montréal pour la majorité des élèves inscrits dans les cours étudiés. SATO-calibrage calcule aussi l'indice de Gunning des textes soumis. En quelques mots, cet indice est une mesure parmi d'autres pour déterminer le niveau de difficulté des textes.

Plus la valeur de l'indice de Gunning est élevée, plus le texte est considéré difficile à lire. Un indice de 6 à 8 correspond à un texte de lecture facile. Un indice de 12 à 15 dénote que le texte est difficile à lire. Bien sûr, il s'agit là d'approximations qui ne considèrent que certains facteurs de lisibilité : la longueur des phrases et la proportion de mots longs (neuf caractères et plus). Plus les phrases sont longues, et plus l'indice de Gunning sera élevé. Une proportion élevée de mots longs va aussi faire augmenter l'indice de Gunning. Il est à noter que l'indice de Gunning a été calibré pour la langue anglaise. (Centre d'analyse de textes par ordinateur, 2007)

Ainsi, il existe une limite à l'utilisation de cet indice en français, mais il peut tout de même contribuer à mieux comprendre l'objet étudié, d'autant plus qu'il est accompagné du détail des informations qui le composent dans la section des résultats.

Le second instrument, VocabProfile, conçu pour l'enseignement des langues secondes, évalue la fréquence du lexique dans la langue française en déterminant le nombre de mots d'un texte

appartenant à la tranche des 1000 mots les plus fréquents de la langue, puis à la tranche des 1001 à 2000 mots les plus fréquents et ainsi de suite. La connaissance acquise sur les caractéristiques des lecteurs grâce au mémoire de Michaud (2014) doit être associée aux résultats obtenus lors de l'analyse des textes par VocabProfile. Le fait de savoir que les élèves ne maîtrisent pas les 1000 mots les plus fréquents doit orienter l'analyse du lexique de ces textes qui servent à mesurer la compréhension en lecture pour des raisons d'équité. En d'autres mots, les textes ne doivent être ni trop difficiles, ni trop faciles — il s'agit de déterminer la zone de développement proximal (Vygotsky, 1978) de ces élèves sur le plan du vocabulaire — et être comparables d'un cours à l'autre.

Il n'est presque jamais utile pour les apprenants de travailler avec des textes comportant plus de 15 % de mots inconnus, à moins que l'objectif soit d'encourager une stratégie d'utilisation constante du dictionnaire ou de deviner (*wild guessing*).

En d'autres termes, nous trouvons le niveau de nos élèves en fonction du palier de 1000 mots qu'ils connaissent, puis nous trouvons ou fabriquons des textes pour eux qui sont à leur niveau actuel ou au niveau suivant selon notre objectif. (Cobb, 2014, TL)

Dans la même veine, Dubay (2004 : 55, TL) indique certaines balises à respecter :

Le problème de la difficulté optimale Différentes utilisations d'un texte nécessitent différents niveaux de difficulté. Comme nous l'avons vu, Bormuth (1969) a indiqué qu'un résultat de 35 % [à 50 %] à un test de closure était le point optimal pour faire des gains sur le plan des apprentissages (...) lors de la lecture assistée en classe. Vygotski (1978) a appuyé les conclusions de Bormuth à l'effet que cette difficulté optimale devrait être légèrement au-dessus de leur niveau de développement et non en dessous. Utiliser des livres qui sont au niveau du lecteur ou au-dessous peut améliorer la fluidité et la vitesse, mais pas la compréhension. Pour cette raison, sachez que le matériel destiné à la lecture assistée quand un enseignant est disponible devrait être un peu plus difficile que le niveau de lecture testé des lecteurs.

L'utilisation du *cloze score* (texte avec des trous tous les cinq mots) est aussi la recommandation de Nation (2009 : 84-85) pour déterminer si un texte est du bon niveau pour les élèves. Si les résultats sont supérieurs à 53 %, le texte est bon pour une lecture sans aide (*independent level*). Si les résultats sont entre 44 % et 53 % (Bormuth dit 35 % à 50 %), le texte est bon pour l'apprentissage (*instructional level*). Si les résultats sont inférieurs à 44 %, le texte est trop difficile (*frustration level*).

Enfin, de futures études portant sur la lisibilité textuelle en FLS pourront sans doute profiter des avancées des travaux de Thomas François (2010, 2014), toujours en cours au moment d'écrire ce rapport.

Tout comme pour l'autre habileté de réception, les évaluations de compréhension de l'oral sont étudiées en fonction de la relation entre la question posée et la réponse dans l'enregistrement audio écouté par les élèves.

3.2. Les habiletés de production : l'expression écrite et orale

Comme dans la section précédente, les habiletés de production sont analysées à partir d'une grille similaire pour chacune. L'objectif est de vérifier si les tâches d'évaluation incluent des éléments communs à ce type de tâche (Legendre, 2005) tels que ceux qui se retrouvent dans le tableau 4 : intention (qui correspond à la contextualisation de Lussier (1992 : 53, 56) : « mise en situation qui laisse croire à une situation de communication la plus authentique possible » avec un but et un destinataire imaginaire), le type de support visuel, l'utilisation ou non de l'ordinateur, le recours à l'évaluation par les pairs ou non, etc.

Tableau 4 Taxonomie pour les évaluations de l'expression écrite

Planification	<ul style="list-style-type: none"> • Présence d'une intention d'écriture (but, contexte) • Anticipation des besoins du destinataire • Choix d'un sujet ou d'un thème • Organisation des idées dans un plan cohérent • Sélection des sources d'information
Rédaction	<ul style="list-style-type: none"> • Sur table • À l'ordinateur
Révision et correction	<ul style="list-style-type: none"> • Stratégies spécifiques identifiées avec l'élève (liste de révision) • Stratégies référentielles en lien avec des règles orthographiques et grammaticales, et non pas intuitives • Temps alloué spécifiquement • Double correction

De la même façon, l'évaluation de l'expression orale a été analysée en fonction des éléments qui se trouvent dans le tableau 5.

Tableau 5 Taxonomie pour les évaluations de l'expression orale

Planification	<ul style="list-style-type: none">• Présence d'une intention (but, contexte)• Anticipation des besoins du destinataire• Choix d'un sujet ou d'un thème• Organisation des idées dans un plan cohérent• Sélection des sources d'information
Présentation	<ul style="list-style-type: none">• Support visuel sur papier• Support visuel électronique• Gestion du temps
Révision et correction	<ul style="list-style-type: none">• Autocorrection considérée dans l'évaluation• Évaluation par les pairs
Langue	<ul style="list-style-type: none">• Qualité de la langue considérée dans l'évaluation

3.3. Définition opérationnelle de l'alignement curriculaire

Enfin, dans le but de préciser l'utilisation du concept d'alignement curriculaire, il est nécessaire de le définir opérationnellement de la façon suivante : dans cette étude, l'alignement curriculaire est une procédure pour mesurer 1) le degré de cohérence verticale entre les objectifs ministériels et les évaluations de plus d'une version d'un même cours ainsi que 2) l'équivalence horizontale des évaluations des différentes versions d'un même cours. D'une part, elle est effectuée par des experts de la discipline qui codent des items d'évaluation en fonction de caractéristiques déterminées et d'autre part au moyen d'instruments informatiques.

L'alignement curriculaire est une procédure pour mesurer 1) le degré de cohérence verticale entre les objectifs ministériels et les évaluations de plus d'une version d'un même cours ainsi que 2) l'équivalence horizontale des évaluations des différentes versions d'un même cours.

4. MÉTHODOLOGIE

Cette recherche est descriptive et quantitative en ce qu'elle vise à établir le degré de cohérence entre les évaluations de quatre cours du niveau 1 en FLS au cégep Vanier, en d'autres mots à

décrire ce degré de cohérence en se référant à différentes mesures prises dans chacune des évaluations des quatre cours étudiés.

Les deux experts de contenu (aussi appelés « experts codeurs » dans la figure 2) ont d'abord collecté les données qui se trouvaient dans les dossiers d'évaluations des quatre collègues ayant accepté de participer volontairement à l'étude. Ces derniers ont été identifiés de P1 à P4. Ces dossiers contenaient toutes les évaluations utilisées dans leur cours de même que les textes à lire et à écouter pour les tâches de compréhension de l'écrit et de l'oral. P1 a aussi soumis des instructions pour aider les experts à comprendre le déroulement d'activités d'interaction à l'oral, une stratégie d'apprentissage qui se retrouve uniquement dans ce cours.

Ensuite, chaque dossier d'évaluations a été numéroté : de 10 à 16 évaluations par cours ont été numérotées, puis l'ensemble des items d'un cours a été numéroté. Il faut distinguer l'item de la tâche d'évaluation. En effet, une tâche donnée peut contenir plusieurs items. En d'autres mots, une question d'examen de lecture ou une tâche d'écriture peut demander à l'élève de faire deux ou trois opérations différentes et chacune recevra un numéro d'« item » afin que chacun soit analysé séparément. Par exemple, si une question de type vrai ou faux demande aussi de justifier la réponse, il y aura deux items pour cette tâche et chacun sera numéroté distinctement afin que le codage des données permette d'affiner l'analyse. Pour revenir à la numérotation des items, c'est donc dire que le premier item de la première évaluation a reçu le numéro 1 et que le premier item de la dernière évaluation (la 10^e, la 12^e ou la 16^e évaluation selon le cas) a pu recevoir le numéro 120, par exemple. Ainsi, la numérotation des items d'évaluation couvre l'ensemble des évaluations du cours.

Une fois les dossiers numérotés, il a été possible de coder les items en fonction des divers facteurs d'analyse choisis (RQR, types de connaissances, etc.). Un premier cours a été codé simultanément par les deux chercheurs afin de partager une compréhension commune des taxonomies. Les trois autres cours l'ont été séparément. Un accord interjuge a été mesuré afin de vérifier le degré de fiabilité du travail effectué. Lorsque le codage n'était pas identique, des choix ont été faits d'un commun accord. Toutes les données ont été saisies dans Excel avant d'être transmises à la statisticienne de l'équipe qui les a traitées avec le logiciel SPSS. Il a été

possible d'attribuer une valeur précise sur le total du cours à chaque item en utilisant une règle de trois simple. Cette opération a permis d'étudier les données sous l'angle du nombre d'items représentant chaque catégorie et de le mettre en contraste avec la valeur réelle que ceux-ci ont dans le cadre du cours.

L'analyse des résultats a été effectuée à l'aide des figures et des tableaux produits lors du traitement des données. Il devenait possible de « lire » les données, d'observer leur représentation et ainsi d'analyser le travail exigé des élèves dans chaque cours.

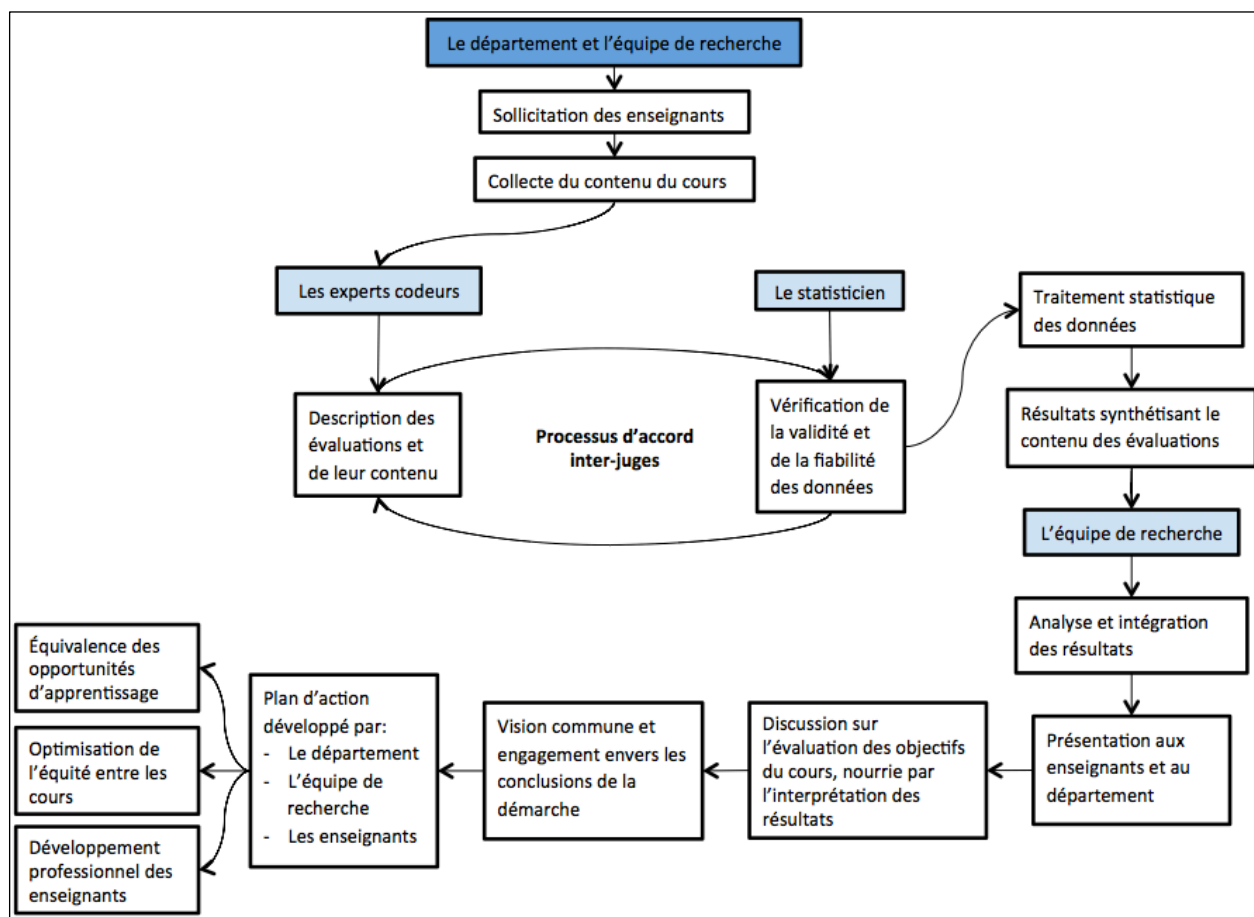


Figure 2 Schéma illustrant le processus d'une analyse d'alignement curriculaire au sein d'un département

L'analyse du degré de lisibilité des textes à lire — aussi appelé le niveau de difficulté — s'est appuyée principalement sur les « caractéristiques superficielles » (Murphy, 2013) des textes, à savoir le nombre de mots par phrases et surtout la fréquence du lexique employé dans la langue française. Tom Cobb, de l'UQAM, a conçu un site Web qui permet de traiter les textes en

français et d'obtenir un rapport détaillé à l'égard de ces facteurs. Cette partie de l'analyse s'est aussi appuyée sur une vue partielle de la connaissance des caractéristiques des lecteurs (le degré de maîtrise des mots les plus fréquents de la langue) et de la tâche de lecture (contexte et modalités de la passation). Une analyse de la connaissance des caractéristiques profondes du texte et des modalités représentées dans le texte (Murphy, 2013) permettrait d'augmenter la fiabilité des tests de lecture d'un cours à l'autre.

L'analyse des résultats a été soumise aux collègues ayant participé à l'étude dans un premier temps sans rétroaction de leur part, puis, par la suite, au département par le biais de ce rapport.

5. RÉSULTATS

Après avoir 1) collecté les dossiers d'évaluation des quatre cours étudiés, 2) encodé les items d'évaluation en fonction des facteurs étudiés dans chaque élément de la compétence, 3) traité les données au moyen de logiciels, il est possible de rendre compte de l'analyse des résultats. Une présentation des données accompagne un commentaire qui vise à décrire l'état de la situation et à souligner les questionnements qu'elles suscitent s'il y a lieu.

5.1. Données descriptives globales

Les données suivantes ne sont pas associées à un élément de la compétence et décrivent les caractéristiques globales de chaque cours. Les données du tableau 6 montrent un certain équilibre entre P1, P3 et P4. Aucune règle ni aucun standard ne concernent le nombre idéal d'items dans un cours, mais il est souhaitable, au nom de la similarité de la charge de travail dans un cours donné par plusieurs collègues, que ce nombre soit similaire entre chacun. P2 pourrait en ajouter une trentaine afin de rejoindre P1. À environ 25 items près entre le cours qui en compte le moins et celui qui en compte le plus, l'équilibre paraît optimal de prime abord, mais il convient aussi d'observer la distribution des items en fonction des objectifs ministériels (voir tableau 7. Le cours 100 est le premier de la séquence du niveau 1 en FLS (formation commune) et le cours HSA est le second (formation propre au programme). Les chiffres entre parenthèses désignent le nombre d'items qui offrent des points bonis aux élèves. Ces points sont accordés en fonction de la complétion de la tâche demandée et non de sa réussite.

Tableau 6 Description du contenu des quatre cours étudiés

Enseignantes ou enseignants	Cours	Nombre d'évaluations (bonus)	Nombre d'items (bonus)	Nombre moyen d'items par évaluation
P1	100	11	175	15,90
P2	100	10 (1)	144 (13)	14,4
P3	HSA	13 (1)	192 (4)	14,77
P4	HSA	17 (1)	198 (10)	11,65

5.2. Description des items

Cette section présente les caractéristiques générales des items, soit celles qui ne se rapportent pas à une taxonomie particulière en éducation. L'objectif ministériel auquel chacun des items se rapporte, ainsi que le type de question et la valeur dans le cours de chaque item sont donc mis en relation.

Le tableau 7 et la figure 3 représentent le nombre d'items que chaque cours comporte pour chacun des objectifs ministériels. Les objectifs 1 et 3 réfèrent aux habiletés productives (écrire et parler) et comptent peu d'items alors que les habiletés réceptives (lire et écouter) associées aux objectifs 2 et 4 en comptent davantage. P1 présente des tâches qui combinent l'écoute et l'expression orale. Les items sans objet (SO) ne correspondent pas à un des quatre objectifs ou font partie des points bonis. Étant donné que les points bonis sont attribués lorsque le travail a été fait sans nécessairement être corrigé comme tel et faire l'objet d'une rétroaction formelle, ces items n'ont pas été codés. La figure 3 ajoute un complément d'information relatif à la proportion des objectifs dans chaque cours. Il ne s'agit pas de la valeur des points attribués à chaque objectif, mais de l'importance relative de chaque objectif dans le cours.

Tableau 7 Nombre d'items par objectif ministériel

Objectifs	P1	P2	P3	P4
1. Rédiger et réviser un texte simple	4	2	3	4
2. Dégager le sens d'un texte simple	60	55	148	127
3. Émettre un message oral simple	1	2	1	3
4. Dégager le sens d'un message oral simple	45	20	35	34
3 et 4	7			
SO	58	65	5	29

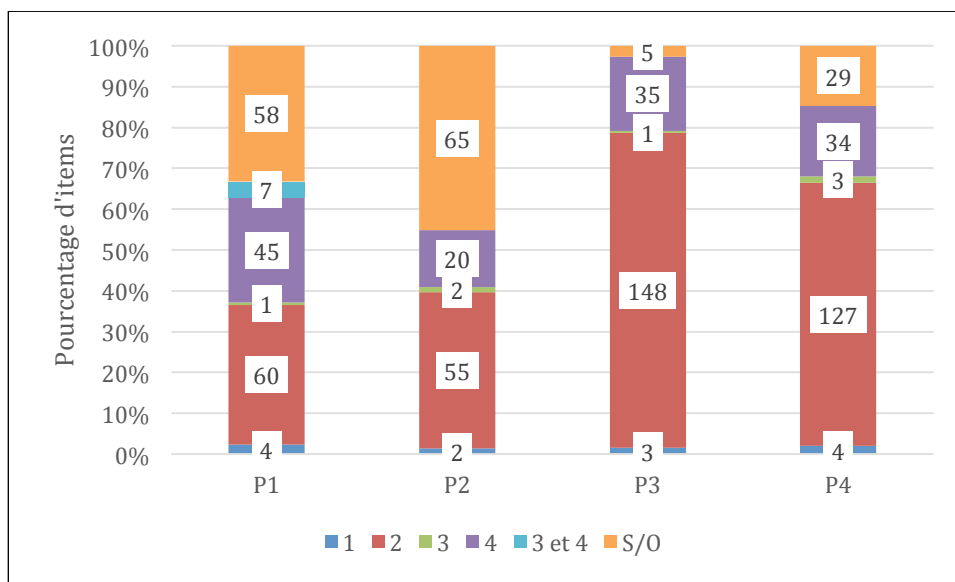


Figure 3 Proportion d'items par objectif ministériel

Les objectifs 2 et 4 (lecture et écoute) présentent des divergences. Les deux cours 100 (P1 et P2) et les deux cours HSA (P3 et P4) sont relativement cohérents en lecture. Il y a le double, voire le triple du nombre d'items dans les cours HSA sans qu'il soit toutefois convenu que cela doive être le cas. En écoute, la divergence se situe dans le cours de P2 qui compte nettement moins d'items en écoute que les 3 autres cours. P1 se distingue aussi des trois autres en offrant des items qui combinent écoute et parole. En somme, le stimulus pour prendre la parole n'est pas lu, mais écouté dans certains cas. Le nombre de performances à l'oral varie chez P4 qui en offre davantage que les autres.

Le tableau 8 et la figure 4 présentent la valeur en points dans le bulletin pour chaque objectif. Les cours de P3 et P4 dépassent 100 % en raison de la présence de points bonis. P2 en a aussi pour 10 % dans le cours, mais ces items n'ont pas été codés parce qu'ils ne sont pas évalués. Les points sont attribués si l'élève fait le travail. De plus, le barème de notation commun n'est pas synonyme d'équivalence entre les cours une fois pris en compte les items sans objet (SO), le nombre d'items et la liberté de chaque collègue à évaluer les travaux et les devoirs qui ne portent pas sur les mêmes objectifs d'un collègue à l'autre. Par exemple, au final l'objectif 2 (lecture) vaut deux fois moins chez P2 que chez P4. L'objectif 3 (expression orale) vaut près de 6 fois plus chez P4 que chez P1, deux fois plus chez P2 que chez P3, etc. L'objectif 4 (écoute) est plus homogène, mais P1 se distingue avec 50 % plus de poids que chez les trois autres et des items où l'écoute et la parole sont combinées.

Tableau 8 Valeur attribuée par objectif ministériel

Objectifs	P1	P2	P3	P4
1	30 %	35 %	40 %	36 %
2	25 %	18 %	35 %	36 %
3	3 %	11 %	5 %	17 %
4	15 %	10 %	10 %	10 %
3 et 4	18 %			
SO	10 %	26 %	14 %	7 %

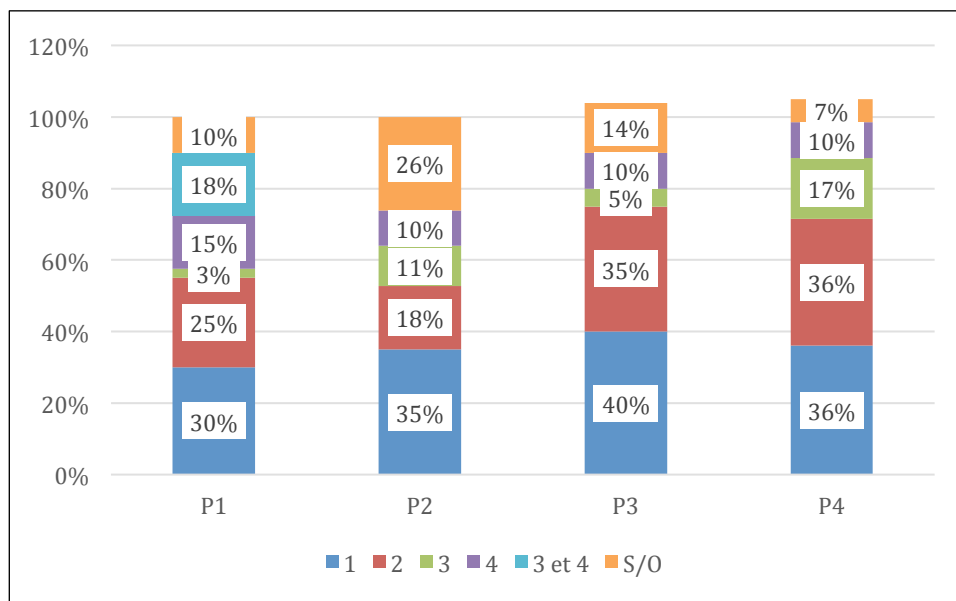


Figure 4 Valeur attribuée par objectif ministériel

Ces données sont importantes, car elles démontrent le fait qu'un barème commun est une mesure insuffisante, voire trompeuse pour assurer l'équivalence de l'évaluation.

Ces données sont importantes, car elles démontrent le fait qu'un barème commun est une mesure insuffisante, voire trompeuse pour assurer l'équivalence de l'évaluation. Comme vu au tableau 2, la lecture des quatre plans de cours indique que tous les membres du département accordent une valeur identique pour l'évaluation de chaque objectif, suggérant qu'il y a équivalence dans l'évaluation. Toutefois,

quand on tient compte du nombre d'évaluations totales (incluant celles de la catégorie discrétionnaire « Contrôles et travaux pratiques ») de même que de leur proportion dans le cours, les écarts apparaissent. Par exemple, la rédaction (objectif 1) ne vaut plus 25 % pour tous les élèves comme dans le plan de cours, mais varie de 30 % à 40 %. L'expression orale (objectif 3) ne vaut pas 10 % pour chaque élève, mais de 3 % à 17 % selon le cours.

P1 a des activités d'apprentissage qui se distinguent des autres cours en ce qu'il ou elle planifie de l'interaction orale où les élèves doivent 1) raconter une histoire sur le modèle proposé par Nation et Newton (2009) appelé le 4/3/2; 2) se poser des questions sur des sujets liés au thème du cours; 3) négocier en équipe de deux pour l'achat d'un produit ou la location d'un appartement, etc. C'est pourquoi certains items ont été codés à la fois pour l'écoute et l'expression orale étant donné que les deux habiletés sont convoquées dans ces activités.

Le tableau 9 permet de discriminer les données par cours et par type de questions posées aux élèves dans les évaluations de chaque cours alors que la figure 5 permet de visualiser les recouvrements et les écarts. Cette dernière est complexe à observer, mais elle recèle une information éclairante à propos des types de questions. Si le développement long (d'une dizaine de lignes environ) présente un certain équilibre, il en va autrement des autres catégories dont certaines se trouvent dans un, deux ou trois cours, mais pas dans les autres. Les deux catégories qui se retrouvent dans tous les cours sont : Développement long et Vrai ou faux non justifié. Les catégories Connaissance du vocabulaire et Reconnaissance du vocabulaire se distinguent de la façon suivante : dans le premier cas, les élèves doivent se souvenir des mots, les connaître dans

la mémoire à long terme alors que dans le second, les élèves se voient présenter une liste de mots qu'ils doivent réutiliser dans un texte à trous, par exemple.

Tableau 9 Nombre d'items par type de question pour chaque objectif ministériel

Objectifs	P1					P2				P3				P4			
	1	2	3	4	¼	1	2	3	4	1	2	3	4	1	2	3	4
Choix Multiples							10		9		29		11		24		6
Dév. Court							12							2	14	1	1
Dév. Long	2					2				3				2			
V/F		22		20			8		11		8		24		11		11
V/F Justifié		22		20			8								7		1
Reconnaissance de vocabulaire							8				50				10		
Connaissance de vocabulaire															5		
Compréhension de consignes écrites				5							1	1					
Grammaire-Vocabulaire	2										5						
Appariement / catégorisation											10				34		
Appariement début-fin phrases											15						
Ordonnancement de phrases											20						
Inférence de vocabulaire		16					9				10						
Oral			1		7			2				1					2
Traduction															3		
Repérage à l'écrit															19		3
Repérage à l'écoute																	12

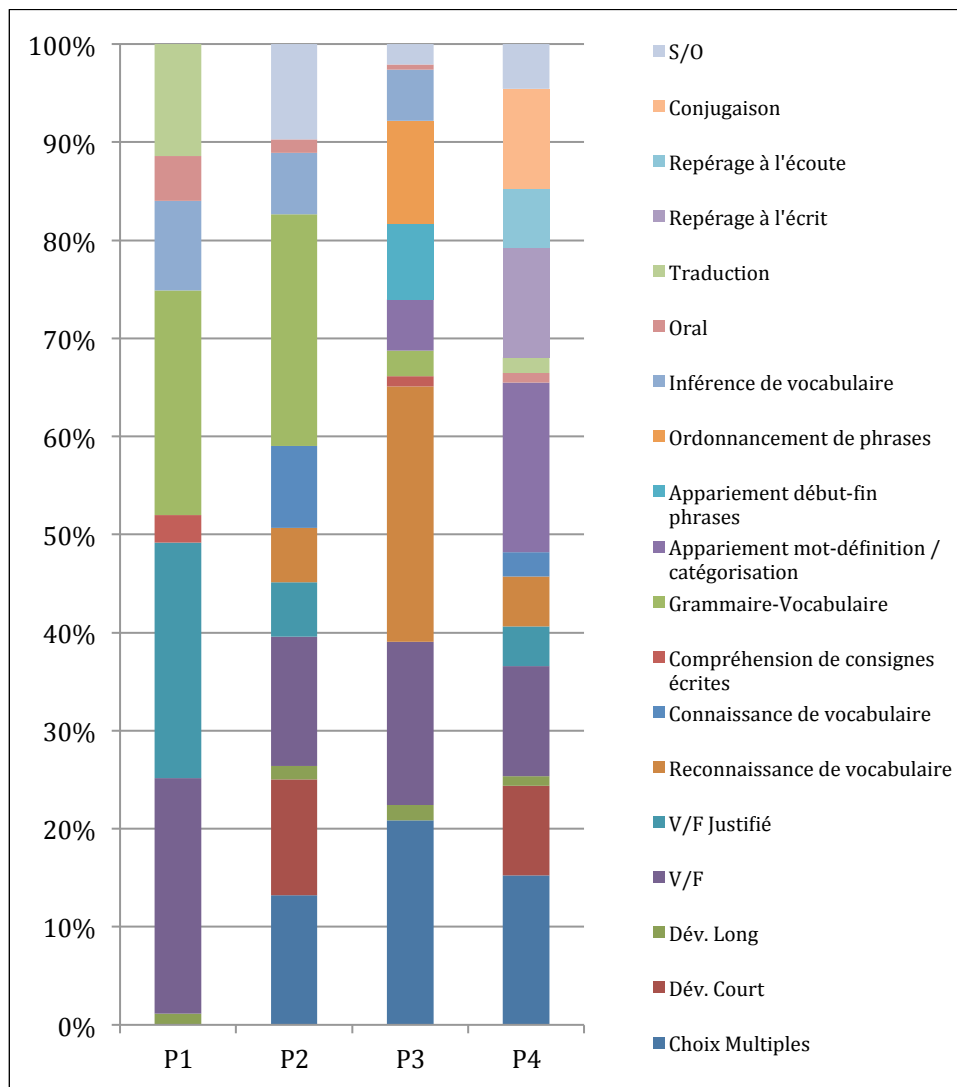


Figure 5 Valeur attribuée à chaque type de question.

Le tableau 10 indique la proportion de chaque format d'évaluation. Ces pourcentages dépendent du nombre total d'évaluations dans le cours. Par exemple, P3 a deux rédactions (EÉ) au programme alors que les trois autres en ont une. Ainsi, P4 semble faire moins de rédactions que P1 et P2, mais comme P4 a 16 évaluations au total et que P2 en a 9, les proportions ne sont pas les mêmes malgré le nombre identique de rédactions. Il est possible de remarquer que P3 ne fait pas d'exposé oral formatif, que P1 et P3 sont les seuls à faire de la compréhension orale (CO) et écrite (CÉ) formatives, que P3 est seul à noter les devoirs et que P4 est seul à noter le plan de l'exposé oral et le plan de la rédaction formative. Il est à noter que les points bonis ont été exclus de cette analyse. Les exercices au laboratoire de compréhension orale et écrite de même que l'exposé oral formatif enregistré au laboratoire sont aussi exclus de l'analyse.

Tableau 10 Proportion d'évaluations selon le format

Format de l'évaluation	P1	P2	P3	P4
1. Quiz – Test	18,18 %	10 %	15,34 %	29,41 %
2. Rédaction	9,09 %	10 %	15,34 %	5,88 %
3. Examen EÉ	9,09 %	10 %	7,69 %	5,88 %
4. Examen CÉ mi-session	9,09 %	10 %	7,69 %	5,88 %
5. Examen CÉ fin de session	9,09 %	10 %	7,69 %	5,88 %
6. Exposé oral formatif	9,09 %	10 %	0,00 %	5,88 %
7. Exposé oral sommatif	9,09 %	10 %	7,69 %	5,88 %
8. Exercice CO+CÉ formatif	9,09 %	0,00 %	0,00 %	0,00 %
9. Examen CO mi-session	9,09 %	10 %	7,69 %	5,88 %
10. Examen CO fin de session	9,09 %	10 %	7,69 %	5,88 %
11. Devoir	0,00 %	0,00 %	15,34 %	0,00 %
12. Plan exposé oral	0,00 %	0,00 %	7,69 %	5,88 %
13. Plan rédaction formative	0,00 %	0,00 %	0,00 %	11,76 %

5.3. La compréhension écrite

Les évaluations de compréhension de l'écrit ont été analysées en établissant la relation entre la question et la réponse (RQR) telle qu'énoncée dans le cadre théorique. Les items sans objet sont des questions de vocabulaire et non de compréhension de texte ou alors des questions de grammaire dans un contrôle où il y a des items de lecture. Ces items sont aussi parfois des questions auxquelles l'élève peut répondre sans avoir lu le texte.

Il n'est pas surprenant de voir au tableau 11 et à la figure 6 qu'il se trouve une majorité d'items se situant dans la catégorie « explicite et textuelle » au niveau 1 en FLS. Les données montrent un certain équilibre qui s'appuie sur la proportion des items dans chaque cours, mais le nombre d'items peut passer du simple au double de P2 à P3. Les questions faisant appel à au moins une inférence, implicites et textuelles, sont équilibrées. Enfin, il conviendrait d'observer attentivement la catégorie « implicite et faisant appel aux schémas du lecteur » et de se demander s'il est pertinent de ne pas avoir d'items de ce type, et plus largement, quelles devraient être les proportions optimales de ces trois types de question à ce niveau en fonction des caractéristiques des lecteurs.

Tableau 11 Nombre d'items par catégorie RQR en compréhension de l'écrit

	P1	P2	P3	P4
Explicite et textuel	44	40	87	53
Implicite et textuel	16	14	20	14
SO	0	1	41	60

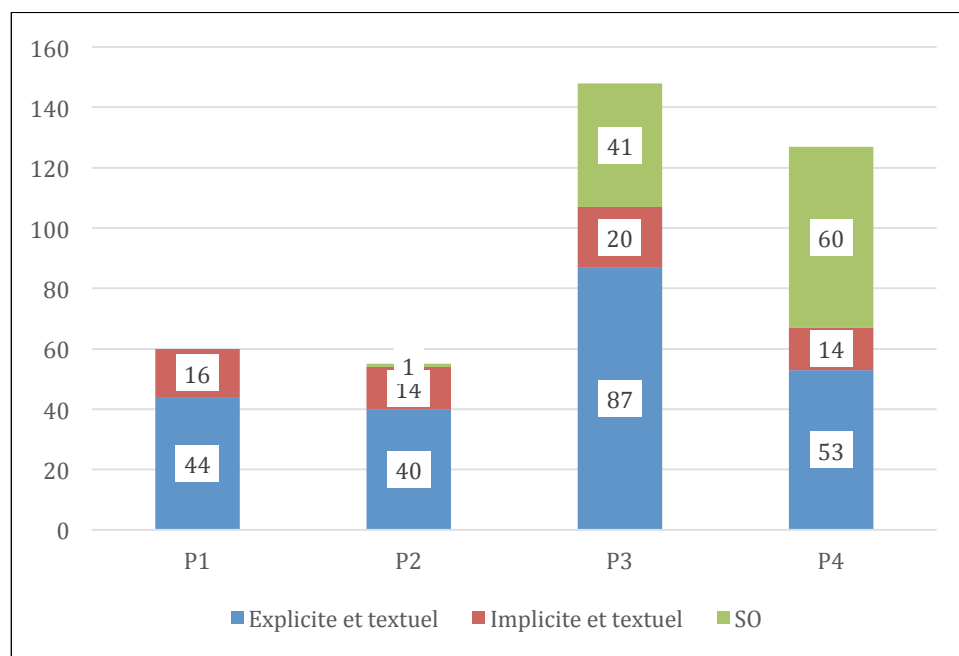


Figure 6 Proportion d'items par catégorie RQR en compréhension de l'écrit

La figure 7 reprend les mêmes données, mais en les présentant sous la perspective de la valeur en points attribuée à chaque catégorie RQR. Ici, l'équilibre est moindre que dans la figure 6, en particulier dans la catégorie « explicite et textuel ». Il est à souligner la valeur relativement importante des items sans objet dans les évaluations de compréhension écrite.

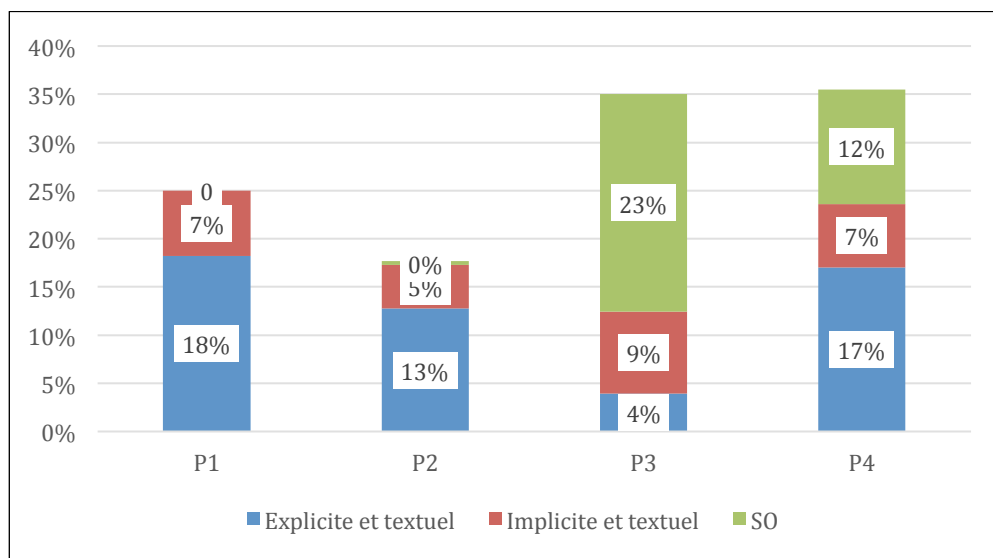


Figure 7 Valeur dans le cours de chaque catégorie RQR en compréhension de l'écrit

5.4. Le niveau de difficulté des textes à lire

Cette partie de l'étude porte sur un élément particulier de l'analyse du degré de difficulté d'un texte, ses « caractéristiques superficielles ». Les quatre sous-sections suivantes présentent des résultats obtenus après avoir traité les textes de chaque examen final de lecture avec SATO-calibrage (Centre d'analyse de textes par ordinateur, 2007; Daoust et al., 1996) et VocabProfile (Selva et collab., 2010), deux instruments informatiques développés à l'UQAM et présentés dans le cadre théorique.

Il s'agit d'une sélection des nombreuses données produites par ces instruments de mesure présentée sous forme de captures d'écran. Le premier tableau, sur fond blanc, montre diverses analyses de la structure du texte de même que l'indice de Gunning correspondant (aussi appelé *Fog Index*) à partir de SATO-calibrage. Le second tableau, sur fond noir, peut être lié à la connaissance du vocabulaire selon la fréquence du lexique dans la langue présentée au tableau 3. VocabProfile mesure la proportion de mots d'un texte qui se retrouvent dans la tranche des 1000 mots les plus fréquents (K-1), puis dans la seconde tranche des 1001 à 2000 mots les plus fréquents (K-2) et ainsi de suite. Après avoir examiné les modalités de passation des examens pour chaque cours de même que les résultats d'analyse avec les deux instruments informatiques, le tableau 12 synthétise les principales données de chaque cours afin d'avoir une vue d'ensemble.

5.4.1. P1

Dans le cours de P1, un nouveau texte doit être lu le jour de l'examen. L'examen, plus court que les autres sur le plan du nombre d'items, est constitué de questions de type vrai ou faux à justifier et d'un exercice d'explication de mots de vocabulaire à partir du contexte. Le tableau sur fond blanc décrit les caractéristiques lexicales et textuelles de façon pointue de même qu'un indice de Gunning de 13.1. Le tableau sur fond noir indique que 71,96 % des mots du texte se trouvent dans la tranche des 1000 mots les plus fréquents, que 78,05 % des mots sont dans la tranche des 2000 mots les plus fréquents (incluant donc la première tranche), et ainsi de suite. La tranche « OFF » contient des néologismes, des mots inconnus ou rares qui ne sont pas dans les 25 000 mots les plus fréquents de la langue.

INDICE GUNNING ET AUTRES STATISTIQUES	
55 mot(s) de 1 car. (10%)	164 mot(s) de 2 car. (29%)
68 mot(s) de 3 car. (12%)	54 mot(s) de 4 car. (10%)
39 mot(s) de 5 car. (7%)	36 mot(s) de 6 car. (6%)
50 mot(s) de 7 car. (9%)	31 mot(s) de 8 car. (5%)
20 mot(s) de 9 car. (4%)	24 mot(s) de 10 car. (4%)
22 mot(s) de 11 - 15 car. (4%)	1 mot(s) de 16 - 20 car. (0%)
0 mot(s) de 21 - 25 car. (0%)	0 mot(s) de 26 - 30 car. (0%)
0 mot(s) de plus de 30 car. (0%)	
nombre de mots..... 564	longueur moyenne: 4.4 car.
nombre de phrases..... 27	longueur moyenne: 20.9 mot(s)
nombre de paragraphes. 1	longueur moyenne: 564.0 mots
pourcentage de mots de 9 lettres et plus: 12%	
indice de lisibilité de Gunning: 13.1	

Current profile (token %)	
K-1 (71.96)	71.96
K-2 (6.09)	78.05
K-3 (10.70)	88.75
K-4 (0.92)	89.67
K-5 (1.85)	91.52
K-7 (0.18)	91.70
K-8 (1.11)	92.81
K-9 (0.37)	93.18
K-10 (0.55)	93.73
K-12 (1.29)	95.02
K-13 (0.37)	95.39
OFF 4.61	100%

5.4.2. P2

Dans le cours de P2, deux textes doivent être lus et annotés avant le jour de l'examen. Des questions portent aussi sur les textes lus et exploités au cours de la session sans que les élèves puissent les consulter pendant l'examen. L'usage du dictionnaire est permis. L'examen est composé d'une question à choix multiple sur le sens général du texte, des vrais ou faux à justifier, des questions de compréhension et des définitions à appairer avec des mots à trouver

dans le texte. Il faut noter que certaines questions qui portent sur les textes lus en classe ne requièrent pas d'avoir lu le texte pour y répondre. De plus, la tâche d'évaluation porte sur la mémorisation d'informations des textes lus au cours de la session au lieu de la démonstration de la capacité à comprendre un texte. Une dernière partie de l'examen porte sur des éléments grammaticaux. Les données présentées ici portent sur les deux textes à lire avant l'examen.

INDICE GUNNING ET AUTRES STATISTIQUES	
107 mot(s) de 1 car. (7%)	353 mot(s) de 2 car. (24%)
246 mot(s) de 3 car. (17%)	162 mot(s) de 4 car. (11%)
107 mot(s) de 5 car. (7%)	124 mot(s) de 6 car. (8%)
87 mot(s) de 7 car. (6%)	102 mot(s) de 8 car. (7%)
74 mot(s) de 9 car. (5%)	38 mot(s) de 10 car. (3%)
70 mot(s) de 11 - 15 car. (5%)	1 mot(s) de 16 - 20 car. (0%)
0 mot(s) de 21 - 25 car. (0%)	0 mot(s) de 26 - 30 car. (0%)
0 mot(s) de plus de 30 car. (0%)	
nombre de mots..... 1471	longueur moyenne: 4.6 car.
nombre de phrases..... 70	longueur moyenne: 21.0 mot(s)
nombre de paragraphes. 6	longueur moyenne: 245.2 mots
pourcentage de mots de 9 lettres et plus: 12%	
indice de lisibilité de Gunning: 13.4	

Current profile (token %)		
K-1	(72.75)	72.75
K-2	(9.84)	82.59
K-3	(4.92)	87.51
K-4	(2.14)	89.65
K-5	(1.57)	91.22
K-6	(0.57)	91.79
K-7	(0.57)	92.36
K-8	(0.93)	93.29
K-9	(0.50)	93.79
K-10	(0.29)	94.08
K-11	(0.43)	94.51
K-12	(0.21)	94.72
K-13	(0.50)	95.22
K-14	(0.07)	95.29
K-15	(0.36)	95.65
K-16	(0.07)	95.72
K-20	(0.07)	95.79
K-21	(0.07)	95.86
K-22	(0.07)	95.93
K-25	(0.07)	96.00
OFF	3.99	100%

5.4.3. P3

Dans le cours de P3, des « tâches de lecture » portent sur des textes lus au cours de la session et deux nouveaux textes. Les « tâches » à propos des textes connus ne sont pas des questions de compréhension ou un rappel d'informations. Les élèves doivent lire et replacer six phrases en

ordre logique. Ils doivent aussi appairer des débuts de phrases avec la fin adéquate de chacune. De plus, ils doivent compléter un texte à trous avec le vocabulaire à utiliser auquel deux mots supplémentaires sont ajoutés, des « distracteurs ». Les tâches portant sur les nouveaux textes — apparentés dans leur thématique avec les textes lus pendant la session —, ne sont pas des questions de compréhension, mais des « tâches de lecture » en soi, sauf une question de compréhension globale du sens du texte. Les élèves doivent appairer le début d'une phrase avec une fin parmi quatre choix proposés, compléter trois phrases à trous sans mots suggérés et inférer le sens de mots à partir du contexte. Il est à noter que l'élève est invité à résumer un texte en une phrase écrite en anglais, mais il n'y a pas de points accordés à cette tâche.

INDICE GUNNING ET AUTRES STATISTIQUES

23 mot(s) de 1 car. (4%)	124 mot(s) de 2 car. (19%)
128 mot(s) de 3 car. (20%)	78 mot(s) de 4 car. (12%)
65 mot(s) de 5 car. (10%)	59 mot(s) de 6 car. (9%)
54 mot(s) de 7 car. (8%)	48 mot(s) de 8 car. (8%)
36 mot(s) de 9 car. (6%)	9 mot(s) de 10 car. (1%)
14 mot(s) de 11 - 15 car. (2%)	0 mot(s) de 16 - 20 car. (0%)
0 mot(s) de 21 - 25 car. (0%)	0 mot(s) de 26 - 30 car. (0%)
0 mot(s) de plus de 30 car. (0%)	

nombre de mots..... 638	longueur moyenne: 4.7 car.
nombre de phrases.... 37	longueur moyenne: 17.2 mot(s)
nombre de paragraphes. 11	longueur moyenne: 58.0 mots

pourcentage de mots de 9 lettres et plus: 9%

indice de lisibilité de Gunning: 10.6

Current profile (token %)	
K-1 (70.09)	70.09
K-2 (8.39)	78.48
K-3 (5.06)	83.54
K-4 (5.22)	88.76
K-5 (1.11)	89.87
K-6 (1.90)	91.77
K-7 (1.27)	93.04
K-8 (0.95)	93.99
K-9 (0.63)	94.62
K-10 (0.16)	94.78
K-11 (0.32)	95.10
K-12 (0.32)	95.42
K-13 (0.47)	95.89
K-14 (0.16)	96.05
K-16 (0.16)	96.21
K-20 (0.32)	96.53
K-21 (0.16)	96.69
K-22 (0.16)	96.85
OFF 3.16	100%

5.4.4. P4

Dans le cours de P4, un nouveau texte doit être lu le jour de l'examen. Les élèves ont accès au dictionnaire et doivent associer des définitions à des mots suggérés. Ils doivent ensuite identifier deux parties de l'introduction, les sujets amenés et posés. L'examen est aussi composé de questions de compréhension où les élèves doivent repérer de l'information, identifier l'idée principale d'un paragraphe, identifier des mots dont le rôle dans le paragraphe est décrit, et expliquer une expression. À la fin du texte à lire, une question d'ordre métacognitif demande à l'élève d'indiquer le degré de difficulté du texte selon lui ou elle et d'expliquer pourquoi. Cette dernière tâche ne compte pas.

INDICE GUNNING ET AUTRES STATISTIQUES			
20 mot(s) de 1 car. (4%)		142 mot(s) de 2 car. (27%)	
99 mot(s) de 3 car. (19%)		57 mot(s) de 4 car. (11%)	
32 mot(s) de 5 car. (6%)		46 mot(s) de 6 car. (9%)	
30 mot(s) de 7 car. (6%)		24 mot(s) de 8 car. (5%)	
21 mot(s) de 9 car. (4%)		23 mot(s) de 10 car. (4%)	
25 mot(s) de 11 - 15 car. (5%)		1 mot(s) de 16 - 20 car. (0%)	
0 mot(s) de 21 - 25 car. (0%)		0 mot(s) de 26 - 30 car. (0%)	
0 mot(s) de plus de 30 car. (0%)			
nombre de mots..... 520		longueur moyenne: 4.6 car.	
nombre de phrases..... 28		longueur moyenne: 18.6 mot(s)	
nombre de paragraphes. 8		longueur moyenne: 65.0 mots	
pourcentage de mots de 9 lettres et plus: 13%			
indice de lisibilité de Gunning: 12.8			

Current profile (token %)		
K-1	(78.67)	78.67
K-2	(8.61)	87.28
K-3	(2.35)	89.63
K-4	(1.17)	90.80
K-5	(3.33)	94.13
K-6	(0.39)	94.52
K-7	(0.78)	95.30
K-8	(0.39)	95.69
K-9	(0.39)	96.08
K-10	(0.39)	96.47
K-11	(0.39)	96.86
K-12	(0.39)	97.25
K-16	(0.20)	97.45
K-18	(0.20)	97.65
K-22	(0.20)	97.85
OFF	2.15	100%

Il a été question dans le cadre théorique du fait que SATO-calibrage s'appuie sur une base de données contenant des textes calibrés en fonction de leur correspondance avec des niveaux scolaires au Québec en langue maternelle. Le tableau 12 présente cet « indice de classement » pour chaque texte traité. Par exemple, les textes lus dans les cours de P1 et P4 correspondent

Le texte de P3 correspond à la 2^e secondaire alors que ceux de P2 sont comparables aux textes lus en 5^e secondaire dans le secteur francophone.

aux textes lus par les élèves de la 3^e secondaire en français langue maternelle au Québec. Le texte de P3 correspond à la 2^e secondaire alors que ceux de P2 sont comparables aux textes lus en 5^e secondaire dans le secteur francophone. Les textes du cours de P2 se démarquent par leur longueur, les phrases en moyenne les plus longues, mais sans l'être

beaucoup plus que les autres. Les textes du cours de P3 sont les plus faciles à tous les égards sauf sur le plan du vocabulaire où c'est l'inverse : les 1000 mots les plus fréquents couvrent 70 % du texte, mais cette donnée est comparable à P1 et à P2. Sur ce plan, le texte de P4 est celui qui présente le lexique le plus accessible à ces élèves.

Tableau 12 Tableau synthèse des caractéristiques superficielles des textes à lire en fin de session au niveau 1

	Nbr. de mots	Nbr. de mots/P	Mots de 9 lettres+	Indice de Gunning	Indice de classement	K-1	K-2	K-5
P1	564	20	12 %	13.1	9.1	71 %	78 %	91 %
P2	1471	21	12 %	13.4	11.2	72 %	82 %	91 %
P3	638	17	9 %	10.6	8.1	70 %	78 %	89 %
P4	520	18	13 %	12.8	9.8	78 %	87 %	94 %

Deux données de ce tableau corroborent le fait que les textes lus sont difficiles. L'indice de Gunning (établi pour des lecteurs de la langue maternelle) est élevé et plus du quart des mots sont au-delà des 1000 mots les plus fréquents qui ne sont pas maîtrisés par les élèves. De plus, Cobb indiquait dans le cadre théorique qu'il était inutile de faire travailler les élèves sur un texte qui comportait plus de 15 % de mots peu ou mal compris. Comme Michaud (2014) a montré que ces élèves connaissent moins de 80 % des mots de la tranche de 1001 à 2000 mots, il y a lieu de se questionner sur le niveau de difficulté des textes au programme. Enfin, le tableau 3 indique que les élèves de ce niveau ignorent 55 % des mots dont la fréquence se situe dans la tranche de 4001 à 5000. De P1 à P3, 10 % du texte est au-delà de cette tranche.

5.5. La compréhension orale

La compréhension de l'orale a été étudiée comme la compréhension écrite, c'est-à-dire en examinant le lien entre la question et la réponse attendue dans les évaluations. Le tableau 13 et

la figure 8 montrent que la presque totalité des items est dans la catégorie « explicite et textuel » sauf chez P4 qui en a deux dans « implicite et textuel ».

Tableau 13 Nombre d'items par catégorie RQR en compréhension de l'oral

	P1	P2	P3	P4
Implicite et Textuel	0	0	0	2
Explicite et Textuel	46	20	35	32
S.O.	6	0	0	0

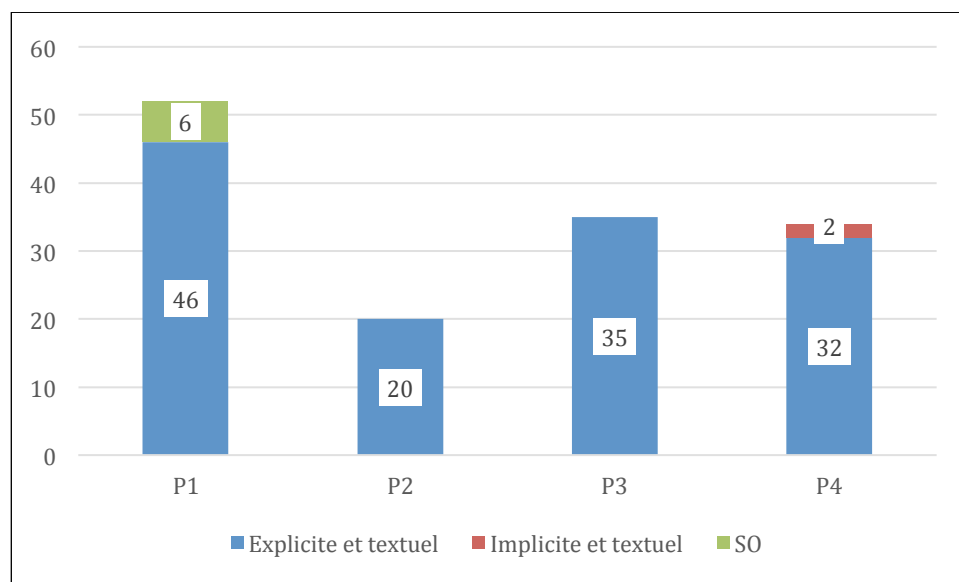


Figure 8 Proportion d'items par catégorie RQR en compréhension de l'oral

Les figures 8 et 9 montrent que dans ce type d'évaluation, l'équilibre est nettement plus manifeste que dans les autres objectifs. La figure 3 a montré que lorsque l'on considère l'ensemble des évaluations des cours, il n'y a pas d'équilibre entre les objectifs sur le plan de leur valeur. Ici, en observant uniquement la compréhension orale, l'équilibre de P2 à P4 s'explique par le fait qu'il n'y a pas d'évaluation de l'écoute

L'équilibre de P2 à P4 s'explique par le fait qu'il n'y a pas d'évaluation de l'écoute dans les Contrôles et travaux pratiques, la catégorie d'évaluations discrétionnaire.

dans les Contrôles et travaux pratiques, la catégorie d'évaluations discrétionnaire. P1 se distingue des trois autres parce qu'il met davantage l'accent sur l'écoute en proposant des

tâches où l'élève doit parler à partir d'un stimulus écouté dans la catégorie Contrôles et travaux pratiques.

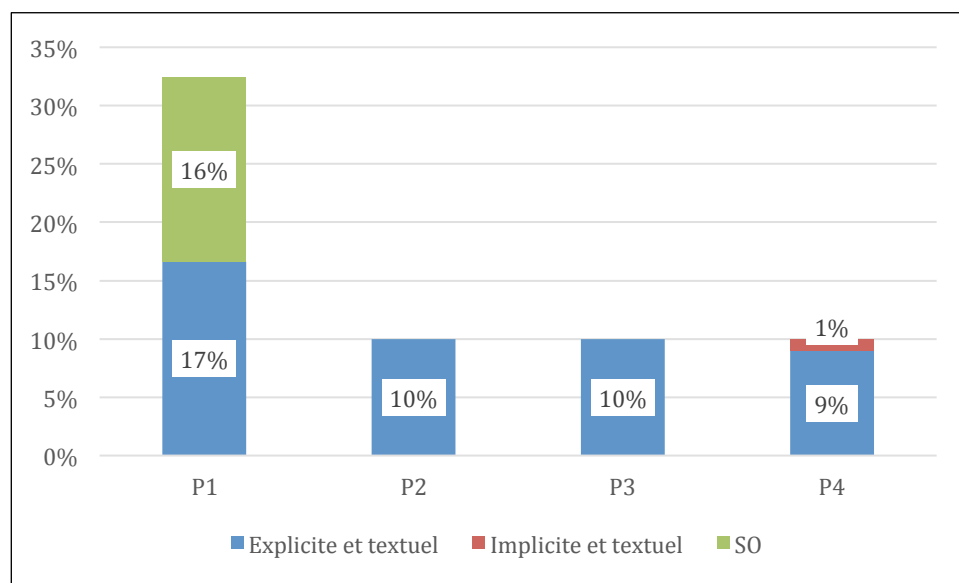


Figure 9 Valeur dans le cours de chaque catégorie RQR en compréhension de l'oral

5.6. L'expression écrite

Les tâches d'évaluation en expression écrite sont distinctes à certains égards, surtout en ce qui concerne la présence d'une intention d'écriture pour l'élève que l'on retrouve seulement chez P2 — expliciter la raison pour laquelle l'élève doit écrire le texte hormis le fait que l'enseignante ou l'enseignant l'exige (figure 10). Cette intention d'écriture est directement liée aux besoins du destinataire dans la tâche que l'on retrouve aussi seulement chez P2. Le plan ne fait pas l'objet d'évaluation chez P1. Aucun des collègues n'évalue l'ajout de sources d'information lors de la rédaction comme telle, ce qui est logique pour P1 et P2 puisque l'élève doit écrire un texte narratif. Toutefois, P3 demande à ses élèves de présenter au moins quatre articles de diverses sources pour la préparation d'un projet de session. Tous les collègues ne font pas rédiger les textes à l'ordinateur, ce qui entraîne des conditions d'évaluation différentes puisque certains ont accès au correcteur informatique (Word et Antidote) et d'autres non. En ce qui concerne les stratégies d'écriture, elles font partie de l'évaluation chez P1, P2 et P3. Il est ici question de stratégies spécifiques identifiées avec l'élève (liste de révision) et de stratégies référentielles (en lien avec des règles orthographiques et grammaticales et non intuitives). Quant à la double

correction, P1, P3 et P4 redonnent à l'élève la moitié des points perdus pour la langue s'il corrige ses fautes. De son côté, P2 code dans un premier temps les erreurs (Vocabulaire = V, etc.) et donne du temps au cours suivant aux élèves afin qu'ils corrigent ces erreurs, voire qu'ils réécrivent certains éléments de fond dans leur texte. P2 corrige ensuite les rédactions et accorde une note finale. Les pratiques à l'égard de la révision ou de la double correction sont donc différentes.

Enfin, il faut mettre en perspective les consignes qui sont accompagnées d'un plan déjà construit pour l'élève où ce dernier doit en quelque sorte remplir les cases. Dans ce cas, il n'a pas à démontrer sa capacité à répondre à la tâche d'écriture en choisissant la structure appropriée. En d'autres mots, il pourrait être placé devant une tâche qui fasse appel à ses connaissances conditionnelles (choisir la structure du plan en fonction de ce qui lui est demandé) et à ses connaissances procédurales (rédaction de la structure choisie).

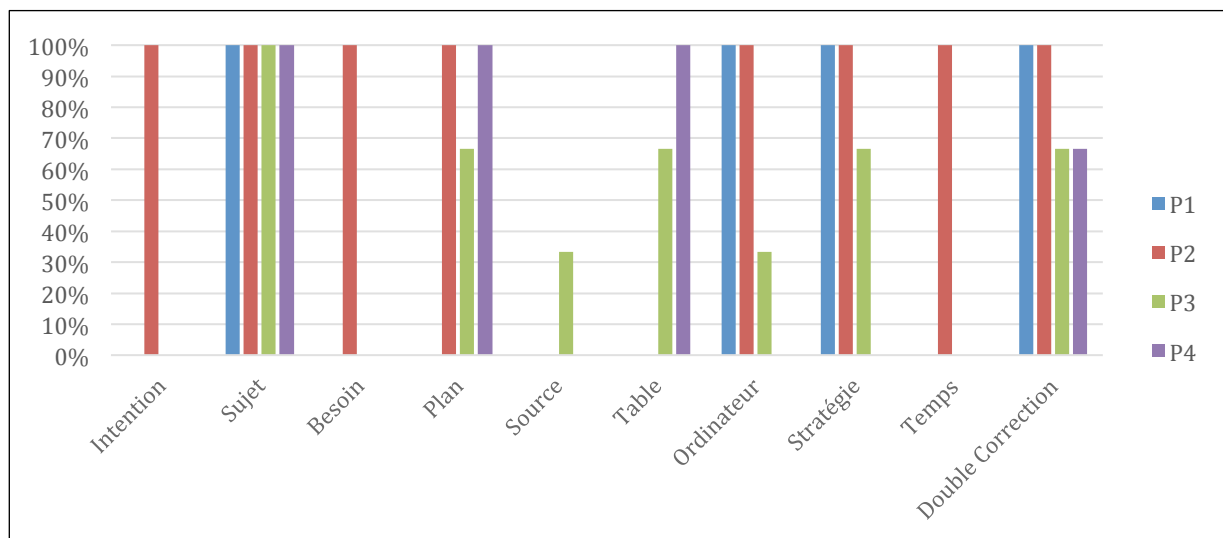


Figure 10 Proportion d'évaluations par catégorie en expression écrite

5.7. L'expression orale

Les évaluations de l'expression orale consistent en un ou deux exposés devant la classe à partir d'un sujet lié à la thématique du cours. P1 et P2 conçoivent des consignes écrites qui présentent à l'élève une intention liée à ce travail en particulier. C'est-à-dire que l'intention n'est pas de faire un exposé oral, mais de négocier la location d'un appartement, de présenter une exposition de photos, en somme une tâche qui mime le réel — ou un contexte authentique —

en tentant de s'éloigner du contexte scolaire et en favorisant l'interaction. La figure 11 montre aussi certaines différences à l'égard du support visuel, sur papier ou sur ordinateur. Dans tous les cours, la gestion du temps lors de l'exposé oral sommatif est évaluée et notée. La figure 11 indique aussi que P4 tient compte de l'autocorrection de l'élève pendant qu'il parle dans l'évaluation. Aucun membre du département participant à l'étude n'inclut dans son évaluation une correction par les pairs d'une façon ou d'une autre. De plus, chez P1, l'évaluation sommative de l'expression orale se fait au moyen d'une production spontanée. C'est-à-dire qu'il ou elle donne le sujet à l'élève quelques minutes avant qu'il ou elle se présente devant le groupe. Cet élève doit 1) raconter pendant quatre minutes la fin d'une histoire qu'il a lue dans le cours ou « vendre » un livre lu pendant le cours lors du programme de *lecture extensive* qui s'appuie sur les travaux de Nation (2009 : 49-60); 2) « réseauter » avec un partenaire devant le groupe pendant 3 minutes; 3) négocier devant le groupe. Des éléments langagiers doivent faire partie de ces performances : vocabulaire, mots interrogatifs, etc. Ces modes d'évaluation se distinguent des pratiques des trois autres cours où l'exposé formel est pratiqué. La qualité de la langue est évaluée dans tous les cours lors de l'épreuve sommative, mais pas de façon formelle dans toutes les activités d'expression orale chez P1, P2 et P4.

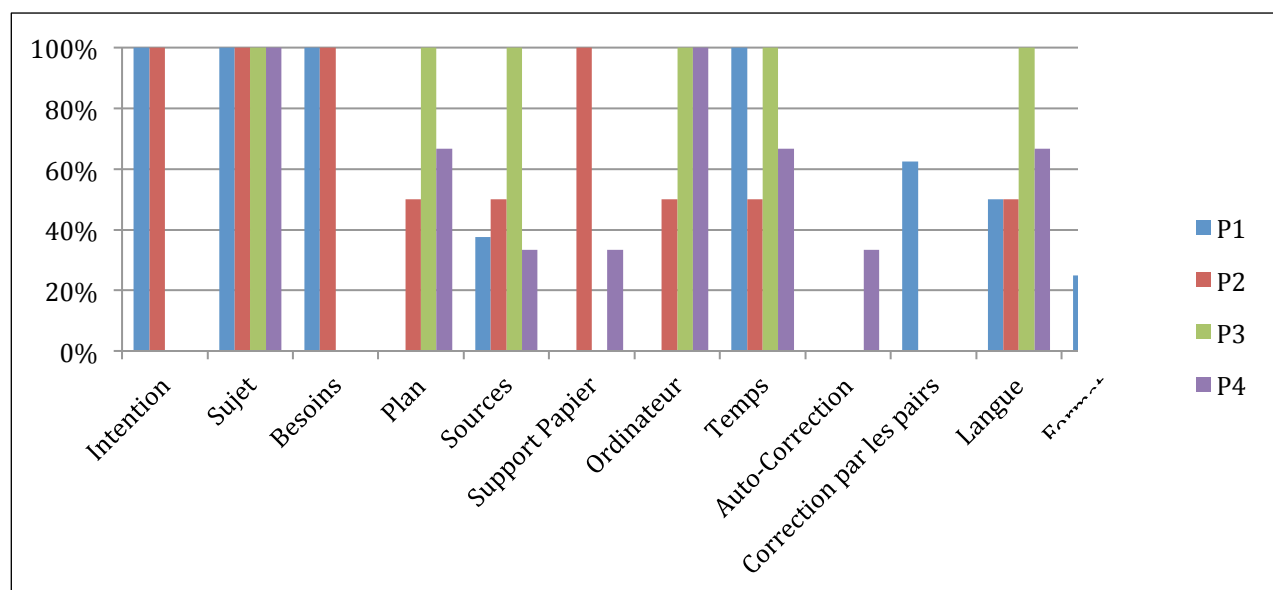


Figure 11 Proportion d'évaluations par catégorie en expression orale

5.8. Les types de connaissances

D'abord, le tableau 14 et la figure 12 montrent que les types de connaissances sont équilibrés pour P3 et P4, mais que P1 et P2 ne le sont pas entre eux ni avec P3 ou P4. S'il est souhaitable d'observer la cohérence entre P3 et P4, il n'en demeure pas moins qu'il est pertinent de se demander quelle devrait être la proportion de chaque type de connaissances à ce niveau. Il peut sembler étonnant qu'un cours de langue ait aussi peu d'items associés aux connaissances procédurales (accords grammaticaux).

Tableau 14 Nombre d'items par type de connaissance

	Déclaratives	Procédurales	Conditionnelles	SO
P1	30	15	130	0
P2	91	28	12	13
P3	70	10	108	4
P4	66	7	113	11

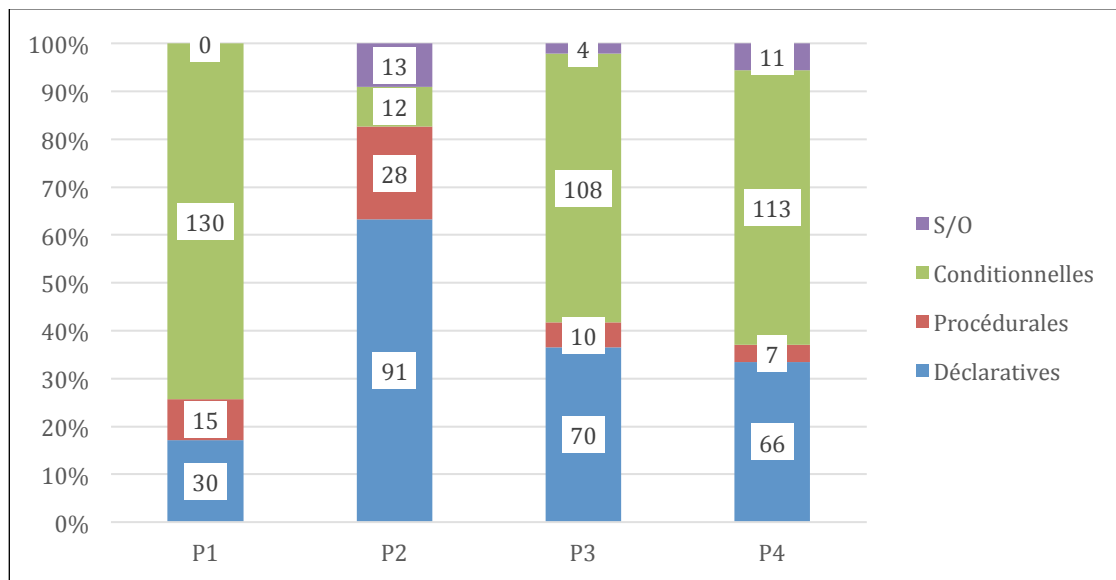


Figure 12 Proportion du nombre d'items selon le type de connaissances

La figure 13 présente les mêmes variables sous l'angle de la valeur attribuée à chaque type de connaissances dans chacun des cours. Les étiquettes dans les colonnes représentent des pourcentages et non des nombres d'items comme dans la figure 12. En comparant ces deux

figures, il est possible de remarquer que les connaissances procédurales chez P1 ont plus d'importance sur le plan de leur valeur que de leur nombre et que c'est l'inverse chez P2. Le constat est opposé chez P2 avec les connaissances conditionnelles qui valent plus que leur nombre contrairement à P1 où on observe l'inverse. Par ailleurs, P3 et P4 conservent un certain équilibre qu'ils partagent avec P1.

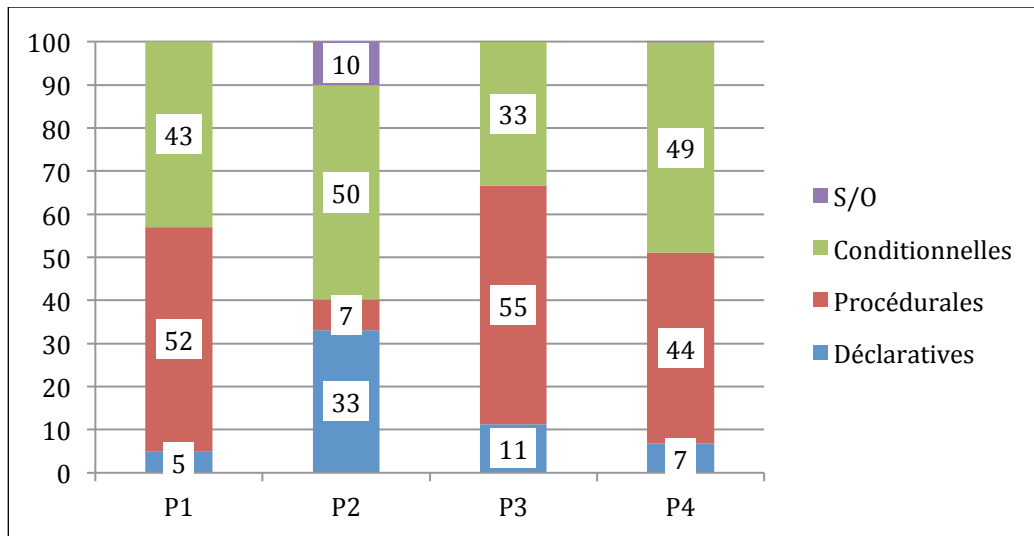


Figure 13 Valeur attribuée par type de connaissances

5.9. Synthèse et conclusion

Après avoir présenté les données descriptives collectées et analysé les évaluations des quatre cours à l'étude en fonction des facteurs choisis, cette partie du rapport vise à rassembler les principaux constats et questions soulevés par l'exercice.

Pour répondre à la question de recherche posée dans la problématique, force est de constater que la réponse n'est pas simple en raison des caractéristiques intrinsèques de la discipline. Chacune des quatre habiletés langagières est très différente des autres et offre de multiples modalités d'évaluation. Toutefois, à la lumière des facteurs analysés pour déterminer le degré d'alignement curriculaire de quatre cours de niveau 1 en FLS au cégep Vanier, il est possible de proposer que des efforts devraient être faits pour augmenter cette cohérence qui apparaît seulement à quelques rares endroits.

Un des éléments importants de cette démonstration est le caractère illusoire d'une cohérence qui s'appuie sur un barème de notation commun. En fait, même s'il n'y avait pas de catégorie discrétionnaire, la variabilité du nombre d'items d'un cours à l'autre opère une distorsion quand les données sont observées sous l'angle de la proportion de chaque catégorie du barème. Qui plus est, la catégorie discrétionnaire est la plus importante du cours avec une valeur de 30 %, ce qui vient accroître l'impact des choix arbitraires qui auront inévitablement une incidence marquée sur la cohérence. Enfin, cette distinction dans le nombre d'items a un impact sur la charge de travail exigée des élèves qui varie d'un cours à l'autre.

Les buts disciplinaires du devis ministériel soulignent la prépondérance de la lecture et de l'écriture dans les cours de FLS au collégial. Il conviendrait de discuter de plusieurs éléments mis en perspective dans l'étude sur ces deux plans.

Ces nombreuses observations et les quelques recommandations tirées des écrits et présentées dans le cadre théorique devraient aider à répondre à plusieurs questions qui se posent, dont les suivantes : le degré de difficulté des textes peut-il être considéré comme « équivalent » au sens de la PIEA? Les modalités de passation

Le degré de difficulté des textes peut-il être considéré comme « équivalent » au sens de la PIEA? Les modalités de passation sont-elles « équivalentes »?

présentées avant les captures d'écran de SATO-calibrage et de VocabProfile sont-elles « équivalentes »? À ce sujet, il appert qu'il y a diverses pratiques dans les cours étudiés où certains testent la lecture en donnant le texte aux élèves avant le jour du test et d'autres le jour même ou alors posent des questions sur des textes travaillés en classe et d'autres non. Ces divergences de modalités d'évaluation sont importantes et devraient faire l'objet de discussions au sein du département. Quel est l'objectif des activités de lecture : connaître le contenu d'un texte en particulier ou démontrer son savoir-faire pour comprendre n'importe quel texte d'un niveau de difficulté similaire à ceux lus en situation d'apprentissage? Quel devrait être le standard pour établir un niveau de difficulté adéquat à ce niveau? L'ensemble des caractéristiques doit-il être pris en considération? Serait-il pertinent de les hiérarchiser? En d'autres mots, quelle pourrait être la meilleure façon de circonscrire la « zone proximale de

développement » en matière de compréhension de lecture, c'est-à-dire une tâche de lecture (niveau de difficulté du texte et types de questions dans un contexte donné) juste assez difficile pour permettre à l'élève de progresser même à travers une évaluation, mais surtout une tâche qui permette de s'assurer que l'élève peut démontrer sa capacité de lecteur à ce niveau? Il conviendrait de déterminer collectivement des critères qui permettraient aux membres du département d'effectuer une sélection de leurs textes et de les adapter au besoin en réduisant le nombre de mots par phrase, en simplifiant le vocabulaire utilisé, en proposant des synonymes ou des périphrases à des mots moins fréquents, voire en traduisant les mots spécialisés peu fréquents, etc. Le recours au test de closure pour valider le choix des textes serait aussi une pratique exemplaire à mettre en place (p. 26 de ce document).

Il serait aussi pertinent de discuter de la présence d'items qui ne sont pas directement liés à la lecture dans des évaluations de compréhension écrite. En effet, la validité de la mesure est compromise puisque le test ne mesure pas uniquement ce qu'il doit mesurer. À ce titre, il serait possible de convenir que la connaissance du vocabulaire est intimement liée à la capacité de comprendre un texte, mais il serait plus difficile d'arriver à la même conclusion avec les notions de grammaire, même s'il s'agit toujours d'une tâche de lecture. En contrepartie, il serait possible d'arguer que de tester la connaissance du vocabulaire appris en classe ne constitue pas d'emblée une mesure de la capacité à comprendre un texte qu'un élève lit pour la première fois.

Un dernier élément lié à la lecture serait l'utilisation du dictionnaire pendant un examen. Nation (2009 : 92) soutient que l'utilisation du dictionnaire ne devrait pas être permise pendant un test parce que les études qui s'intéressent aux facteurs liés à la lecture montrent que la connaissance du vocabulaire est la principale composante de la compréhension en lecture. Il suggère d'utiliser des textes adaptés au niveau de l'élève afin que ce dernier puisse démontrer l'utilisation des stratégies de lecture. Il ajoute que si l'objectif du test est de savoir si l'élève est capable de lire un texte difficile avec soutien, alors les dictionnaires pourraient être permis.

Des questions similaires se posent pour chaque élément de la compétence et en particulier pour l'expression orale au niveau 1, un niveau où l'élève ne peut s'exprimer sans avoir recours à

l'anglais dans presque toutes ses phrases. En particulier, un programme d'expression orale par compétences devrait offrir plusieurs mesures afin que l'élève ait l'occasion d'apprendre de ses erreurs en contexte purement formatif sans perdre de points. Les cours où les élèves ont des exercices d'expression orale au laboratoire de langues s'inscrivent dans cette perspective.

En écriture, les cours étudiés n'ont pas de divergences cruciales, mais il serait intéressant de discuter de la conception de tâches qui permettent l'établissement d'une intention d'écriture pour l'élève.

En écoute, les divergences sont aussi moins frappantes. Ces deux dernières habiletés langagières sont celles qui comptent le moins d'items, réduisant ainsi les risques d'écarts. Il y aurait tout de même lieu de se questionner sur les types de connaissances invoquées dans ces tâches d'évaluation.

Plus globalement, une autre distinction importante qui se dégage de cette analyse est l'attribution ou non de points bonis. Cette attribution a une incidence sur la réussite et peut invalider la note de l'élève. En effet, un résultat de 50 % est un échec dans un cours sans points bonis et pourrait être une réussite dans un cours avec points bonis.

En terminant, il est pertinent de rappeler que l'équipe départementale a décidé de suspendre les analyses d'alignement curriculaire pendant la réalisation de cette seconde étude. Cette décision est légitime, mais ne fait pas disparaître les responsabilités qui incombent au département par le biais de la PIEA du Collège. Il faut aussi rappeler que ces discussions, qui reviendront notamment pour ce qui a trait à l'équité envers les étudiants sur le plan des pratiques d'évaluations, gagnent à s'appuyer sur des faits objectifs, mesurables, fiables et valides. Dans le cas contraire, la subjectivité (« Moi, dans mon cours... ») et le relativisme (« Chacun est libre de faire ce qu'il veut parce qu'on est des professionnels ») nuisent à l'atteinte de l'objectif qui consiste à favoriser des pratiques d'évaluation équitables pour les élèves.

RÉFÉRENCES

- Anderson, L. et Krathwohl, D. (2001). *A Taxonomy for Learning, Teaching, and Assessing*. Longman New York. Longman. Récupéré du site : <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:A+taxonomy+for+Learning,+teaching,+and+assessing#1>
- Bateman, D., Taylor, S., Janik, E. et Logan, A. (2009). Curriculum coherence and student success. *Pédagogie Collégiale*, 22(5), 8–18.
- Beaumier, J.-P. (2012). *L'application des politiques d'évaluation des apprentissages. Pour des évaluations justes et équitables*. Québec: Commission d'évaluation de l'enseignement collégial (Gouvernement du Québec). Récupéré du site : http://www.ceec.gouv.qc.ca/publications/SYNTHESES/ApplicationPIEA_synthese2012.pdf
- Brennan, R. L. (dir.). (2006). *Educational measurement* (4^e éd.). Westport: American Council on Education and Praeger Publishers. Récupéré du site : <http://www.loc.gov/catdir/toc/ecip0614/2006015706.html>
- CEEC. (2002). *Politique institutionnelle d'évaluation des apprentissages du Cégep Vanier. Deuxième rapport d'évaluation*. (J. L'Écuyer, dir.). Montréal: Commission d'évaluation de l'enseignement collégial (Gouvernement du Québec). Récupéré du site : <http://www.ceec.gouv.qc.ca/publications/PIEA/Vanier2.pdf>
- CEEC. (2006). *Évaluation du programme Science (200.B0) conduisant au diplôme d'études collégiales (DEC) au Collège régional Champlain — Saint-Lambert*. (Nicole Lafleur, dir.). Montréal: Commission d'évaluation de l'enseignement collégial (Gouvernement du Québec). Récupéré du site : http://www.ceec.gouv.qc.ca/publications/Programmes_Operation2005/Chamlambert.pdf
- CEEC. (2011). *Évaluation de l'application de la politique institutionnelle d'évaluation des apprentissages du Collège Vanier*. Montréal.
- Centre d'analyse de textes par ordinateur. (2007). Manuel de référence. Récupéré du site : http://www.ling.uqam.ca/sato/satoman-fr_analyseur_lisibilite.html
- Cobb, T. (2014). Matching learners to texts - a rough guide [in “click here to see how”]. Récupéré du site : <http://www.lexutor.ca/vp/>
- Daoust, F., Laroche, L. et Ouellet, L. (1996). SATO-CALIBRAGE: présentation d'un outil d'assistance au choix et à la rédaction de textes pour l'enseignement. *Revue Québécoise de Linguistique*, 25(1). Récupéré du site : <http://www.ling.uqam.ca/sato/index.html>

- Dubay, W. H. (2004). *The Principles of Readability*. Récupéré du site : <http://www.nald.ca/library/research/readab/readab.pdf>
- François, T. (2010). La lisibilité computationnelle. *International Journal of Applied Linguistics*, 160, 75–99. doi:10.2143/ITL.160.0.2129558
- François, T. (2014). Dmesure : lisibilité des exercices de FLE. Récupéré du site : <http://cental.fltr.ucl.ac.be/projects/complexites/>
- Gagné, Ph., Brunet, S., Boucher, G. et Dumont, L. (2012). *Alignement curriculaire en français langue seconde au cégep Vanier. Cours de niveau 3*. Montréal: Vanier College Press. Récupéré du site : <http://www.cdc.qc.ca/pdf/788353-gagne-et-al-alignement-curriculaire-FLS-vanier-2012.pdf>
- Gauthier, C., Bissonnette, S., Richard, M. et Castonguay, M. (2013). *Enseignement explicite et réussite des élèves. La gestion des apprentissages*. Saint-Laurent: ERPI.
- Gauthier, C., Mellouki, M., Bissonnette, S. et Richard, M. (2005). Écoles efficaces et réussite scolaire des élèves à risque. Un état de la recherche. Québec: Université Laval.
- Germain, L. (germainl@csgm.qc.ca). *Validité et fidélité de l'évaluation*, [Message électronique à Philippe Gagné], (gagnep@vaniercollege.qc.ca), 1^{er} juin 2015.
- Giasson, J. (2011). *La compréhension en lecture* (3^e édition.). Bruxelles: De Boeck.
- Irwin, J. (1986). *Teaching Reading Comprehension Processes*. Englewood, New Jersey: Prentice-Hall.
- Jeffrey, D. (2009). L'éthique dans l'évaluation scolaire. *Formation et Profession*, 16(1), 47–49. Récupéré du site : <http://www.crifpe.ca/download/verify/495>
- Koretz, D. M. et Hamilton, L. S. (2006). Testing for Accountability in K-12. In R. L. Brennan (dir.), *Educational Measurement* (4^e éd., pp. 531–578). Westport: American Council on Education and Praeger Publishers.
- Legendre, R. (2005). *Le dictionnaire actuel de l'éducation*. Montréal: Guérin.
- Lussier, D. (1992). *Évaluer les apprentissages dans une approche communicative*. (S. Moirand, dir.). Paris: Hachette.
- Martone, A. et Sireci, S. G. (2009). Evaluating Alignment Between Curriculum, Assessment, and Instruction. *Review of Educational Research*, 79(4), 1332–1361. doi:10.3102/0034654309341375

- MELS. (2011). *General Education. Common, Specific, and Complementary to Programs Leading to a Diploma of College Studies*. (Direction de l'enseignement collégial, dir.). Québec: Gouvernement du Québec. Récupéré du site : http://www.mesrst.gouv.qc.ca/fileadmin/contenu/documents_soutien/Ens_Sup/Collegial/Form_collegiale/FormGenComPropreComplProgEtudesCondDEC_2011_1_a.pdf
- Michaud, G. (2014). *Relations entre morphologie, vocabulaire et compréhension écrite en français langue seconde* (Mémoire de maîtrise non publié). Université de Montréal, Montréal.
- Murphy, S. (2013). Évaluer la difficulté des textes pour les élèves. Récupéré du site : http://www.edu.gov.on.ca/fre/literacynumeracy/inspire/research/WW_ATDSFr.pdf
- Nation, I. S. P. (2009). *Teaching ESL/EFL Reading and Writing*. New York: Routledge.
- Nation, I. S. P. et Newton, J. (2009). *Teaching ESL/EFL Listening and Speaking*. New York: Routledge.
- Organisation de coopération et de développement économiques (OCDE). (2012). *Literacy, Numeracy and Problem Solving in Technology-Rich Environments: Framework for the OECD Survey of Adult Skills*. Éditions OCDE. Récupéré du site : [//dx.doi.org/10.1787/9789264128859-en](http://dx.doi.org/10.1787/9789264128859-en)
- Pearson, D. et Johnson, D. (1978). *Teaching Reading Comprehension*. New York: Holt.
- Ravitch, D. (2010). *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*. Philadelphie: Basic Books.
- Roegiers, X. (2010). *L'école et l'évaluation. Des situations complexes pour évaluer les acquis des élèves*. (J.-M. De Ketele, dir.) (2^e éd.). Bruxelles: De Boeck.
- Sabourin, Y. (2009). Un regard sur le département. FNEEQ-CSN. Récupéré du site : <http://www.fneeq.qc.ca/fr/cegep/outils/Formation-departements-FR-2011-04.pdf>
- Selva, T., Jones, G., Goodfellow, R. et Cobb, T. (2010). Web VocabProfil en français. Récupéré du site : <http://www.lex Tutor.ca/vp/>
- Squires, D. A. (2009). *Curriculum Alignment. Research-Based Strategies for Increasing Student Achievement*. Thousand Oaks: Corwin Press.
- Tardif, J. (1997). *Pour un enseignement stratégique. L'apport de la psychologie cognitive*. Montréal: Les éditions Logiques.

Vanier College. (2011). *Institutional Policy on the Evaluation of Student Achievement (IPESA)*. (Direction des études, dir.). Montréal: Vanier College. Récupéré du site : http://www.vaniercollege.qc.ca/policies/uploads/7210-38_IPESA-1.pdf

Vygotsky, L. (1978). *Mind in Society*. (E. S. Cole, M., V. John-Steiner, S. Scribner, dir.). Cambridge, MA: Harvard University Press.

Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education (Research Monograph No. 6)*. Washington, DC.: Council of Chief State School Officers.